

THIS PAGE BLANK (USPTO)

DB 557 TSPKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 595

RESULT 2

Q9SJ06 PRELIMINARY; PRT: 1207 AA.

ID Q9SJ06

AC Q9SJ06

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 19, Last annotation update)

DE At2g36490 protein.

GN At2g36490

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RC MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

RT thaliana.";

RT Nature 402:761-768 (1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X.;

RA Submited (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC006919; AAD24633.1;

DR InterPro: IPR003651; Fes_bind.

DR SMART: SM00525; FES; 1.

QY SEQUENCE 1207 AA; AADB6C9EE54DDF67 CRC64;

Query Match 59.5%; Score 163.5; DB 10; Length 1207;

Best Local Similarity 43.4%; Pred. No. 9.7e-20;

Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;

QY 1 KVXXDXXTXXWXXL-----XXXDXXXXXXXXERXXFXRXXMXXXGXGR 52

DB 500 VLDLDETRVWKLLENINSEGVDSDEQAKWEEERNVFRGRADSFARMHLVQGR 559

QY 53 FXFXKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 595

DB 560 RTPKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 598

RESULT 3

O49498 PRELIMINARY; PRT: 917 AA.

AC O49498

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE Hypothetical 106.3 kDa protein.

GN F28A23.180 OR A74G34060.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

RT thaliana.";

RT Nature 402:761-768 (1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X.;

RA Submited (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC006919; AAD24633.1;

DR InterPro: IPR003651; Fes_bind.

DR SMART: SM00525; FES; 1.

QY SEQUENCE 1207 AA; AADB6C9EE54DDF67 CRC64;

Query Match 59.5%; Score 163.5; DB 10; Length 1207;

Best Local Similarity 43.4%; Pred. No. 9.7e-20;

Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;

QY 1 KVXXDXXTXXWXXL-----XXXDXXXXXXXXERXXFXRXXMXXXGXGR 52

DB 500 VLDLDETRVWKLLENINSEGVDSDEQAKWEEERNVFRGRADSFARMHLVQGR 559

QY 53 FXFXKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 595

DB 560 RTPKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 598

RA Meves H.W., Mayer K., Schueller C.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,

RA Neumann S., Meves H.W., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL021961; CAAL7566.1;

DR EMBL: AL161584; CAB80123.1;

DR InterPro: IPR003651; Fes_bind.

DR InterPro: IPR003651; Fes_bind.

DR Pfam: PF00730; HHG-GPD; 1.

DR SMART: SM00478; ENDO3C; 1.

DR SMART: SM00525; FES; 1.

KW Hypothetical protein.

QY SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match 58.9%; Score 162; DB 10; Length 917;

Best Local Similarity 43.9%; Pred. No. 1.3e-19;

Matches 43; Conservative 0; Mismatches 47; Indels 8; Gaps

QY 1 KVXXDXXTXXWXXL-----XXXDXXXXXXXXERXXFXRXXMXXXGXGR 53

DB 418 KYNLPETIKEDVLMVNDSPSRVDDTEAKWKEIEFQTRIDLFNRHRLGNRK 477

QY 54 FXFXKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 90

DB 478 FKQKMGVSDVSVGVFLTQNVADHSSSSAAMDAAEFP 515

RESULT 4

O9LZ69 PRELIMINARY; PRT: 1017 AA.

ID O9LZ69

AC O9LZ69

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE Hypothetical 112.1 kDa protein.

GN T32M21.160.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Teiryn N., Ardiles W., Buyschaert C., Dasseville R.,

RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,

RA Villarroel R., Gielen J., Van Montagu M., Bancroft I., Meves H.W.,

RA Rudd S., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL162875; CAB85562.1;

KW Hypothetical protein.

QY SEQUENCE 1017 AA; 112139 MW; 4F7C313A891EC83F CRC64;

Query Match 18.0%; Score 49.5; DB 10; Length 1017;

Best Local Similarity 28.8%; Pred. No. 9.4;

Matches 17; Conservative 0; Mismatches 33; Indels 9; Gaps

QY 1 KVXXDXXTXXWXXL-----XXXDXXXXXXXXERXXFXRXXMXXXGXGR 50

DB 931 KVIDDETTIWNLMGKDEKDKKKEWEEERVFGRADSFARMHLVQGR 989

RESULT 5

THIS PAGE BLANK (uspto)

A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: D84198
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <STO>
A:Cross-references: GB:AE004437; NID:g10580010; PIDN:AAG18952.1; GSPDB:GN00138
C:Genetics:
A:Gene: nthA1

	Query Match	12.9%	Score 89.5;	DB 2;	Length 265;
	Best Local Similarity	27.6%;	Pred. NO. 3.9e-06;		
	Matches 3;	Conservative	8; Mismatches 53;	Indels 23;	Gaps 5;
Qy	88 KXXLLXXGXGLKXSECVRLLLXX - XAFPVDTNWGRIXVRXGVPLXP L P L P X X Q X H X L X	146	:	:	:
Dd	145 RAALLDNTGVPKTADCVLLFAGGAPGVDPDVTHVRIARMGMIAAADHEAVRAALEA	204	:	:	:
Qy	147 XYPXXXXXQLPRLCKLXOXTLYELHYXITTEGKXXFCFKXKPNC ----NACPM	198	:	:	:
Dd	205 AVPDAA-----CGEG-----HTAMIOEGR-EYCTARPDPACILDDPAACP	242	:	:	:

RESULT 7
G87530
endonuclease III family protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: G87530
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: AB7249; MUID:21173698; PMID:11259647

A:Accession: G87530
A:status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 S>TO>
A:Cross-references: GB:AE005673; NID:gl3423785; PIDN:AAK24243.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2272

```

Query Match      12.5%; Score 86.5; DB 2; Length 241;
Best Local Similarity 24.2%; Pred. No. 1.le-05;
Matches 31; Conservative 9; Mismatches 73; Indels 15; Gaps 3;

Qy 72 GXIDLEWLRXXXXXXXXLLXXGXGLKSECV-RLLLXLLXXAFPVDYTNVGRIXVRXGX 130
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 106 GWLSLSHLKLTVDQARWELQALPGVGKVAACVNFSDLAMRALVWDTHVDRVARRIGL 165
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 131 VPLPLPXXXXXHXLLXXYPXXXXXXOKXLLWPLRCLXQXTYELHYXXITFGKXXFXCTXX 190
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 V-----GSGDTNTNYHTLMAAPASW-----TADDLFELHWMKRGQLCGAEA 211
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 191 PNCNACPM 198
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 212 PKCGACPV 219
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8
A75109
endonuclease III (nth) PAB0459 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: A75109
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: A75109
A>Status: preliminary
A:Molecule type: DNA

A: Residues: 1-222 <KAW>
 C: Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49586.1; PID:g5454
 C: Experimental source: strain Orsay
 C: Genetical
 C: Gene: nfh; PAB0459
 C: Superfamily: apurinic/apyrimidinic endonuclease III
 C: Keywords: 4Fe-4S; metalloprotein
 E: 195,202,205,211/Binding site: 4Fe-4S cluster (Cys) #status predicted

	Query Match	12.4%;	Score 86;	DB 2;	Length 222;
	Best Local Similarity	26.9%;	Pred. No. 1.2e-05;		
	Matches	29;	Conservative 11;	Mismatches 50;	Indels 18; Gaps
Qy	91	LLXXXGXGLKXECVRLLLX	XXAxFPVD	NVGRGXVXR	GXVPLXPLPXXXXXHXHLLXXYPX 150
Db	118	LILPGIGIRKCANIVLAYGFI	PAIPVD	THVYISRLGLAP	WDASPEEVEERL----- 171
Qy	151	XXXXQKLWPLRCLKQXTLYE	LHYXXITF	GKXXFXCTK	XXPNACNACPM 198
Db	172	-----KELIPR-----E	EVIY-VNHAMVDHGK-	SVCRP	IKPRCDECP 207

RESULT 9
B69202
endonuclease III - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
C:Accession: B69202
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
Q; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
k. I. S.; Church, G.M.; Daniels, C.-J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000: MUID:98037514; PMID:93711463

A:Accession: B69202
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <MTH>
A:Cross-references: GB:AE000855; GB:AE000666; NID:g2621852; PIDN:AAB85267.1; PID:g362
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH764
A:Start codon: TTG
C:Superfamily: apurinic/aprimidinic endonuclease III

Query Match	12.08;	Score	83.5;	DB 2;	Length	233;
Best Local Similarity	26.99;	Pred. No.	3.le-08;			
Matches	29;	Conservative	10;	Mismatches	50;	Indels
						19;
						Gaps
Qy	91	LLXXGXGLKSEXCEVRLRLX	XXXXAPPVDTN	VGRIXVRXGVPLPL	PLPXXXXXXHLXXVFX	150
Db	127	LLKLPVGRKTCANGLVYAFGR	PAIPVDTVHRISSNRIGLVD	-TRTPEETERALMKVIP-	184	
Qy	151	XXXXOKLWPLRCLKXQXTLY	ELAHVXXITFGKXXFCTKXX	PNCNACPM	198	
Db	185	-----RWV-----	-----TELNDLMVQFOG	-DICRPLGPRHRECP	215	

RESULT 10
F71025
probable endonuclease III - Pyrococcus horikoshii
C/Species: Pyrococcus horikoshii
C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C/Accession: F71025
R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A/Reference number: A71000; MUID:98344137; PMID:9679194
A/Accession: F71025
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-222 <RAW>

THIS PAGE BLANK (uspto)

THIS PAGE BLANK (uspto)

A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ss

A:Reference number: A86625; MUID:21235186; PMID:1137471

A:Accession: G86758

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-218 <STO>

A:Cross-references: GB:AE005176; PID:g12724026; PIDN:AAK05169.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: nth

C:Superfamily: apurinic/apryrimidinic endonuclease III

C:Keywords: carbon-oxygen lyase

Query Match 10.7%; Score 74; DB 2; Length 218;

Best Local Similarity 25.9%; Pred No. 0.00092;

Matches 30; Conservative 9; Mismatches 49; Indels 28; Gaps 4;

QY 88 KXXLLXXGXGKXECVRLXLLXXXAFVDTNVGRXVPLXPLPXXXXHXLXX 147

Db 112 KKVQLTLPVGRKTNVLAENVGIPGIAVDTHVERVSKRLDIVP----- 156

QY 148 YPXXXXXX-----LWPRCLKX-QXTLYELHYXXITFGKXXFCYXXPNCNACPM 198

Db 157 -----QKATVLEVEEKMLIPOEKVQAOHHHLIFFGR-YHCTAKPKCADCPV 204

RESULT 15

AF1311

probable endonuclease III (DNA repair) [imported] - *Listeria monocytogenes* (strain EGD-e

C:Species: *Listeria monocytogenes*

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AF1311

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant,

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: ABI077; MUID:21537279; PMID:11679669

A:Accession: AF1311

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-219 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAC99972.1; PID:g16411347; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: nth

C:Superfamily: apurinic/apryrimidinic endonuclease III

Query Match 10.4%; Score 72; DB 2; Length 219;

Best Local Similarity 25.2%; Pred No. 0.0019;

Matches 26; Conservative 8; Mismatches 51; Indels 18; Gaps 3;

QY 96 GXGLKXECVRLXLLXXXAFVDTNVGRXVPLXPLPXXXXHXLXXYPXXXXXQ 155

Db 117 GVGKRTANVLSVGFVPAIAVDTHVERISKRLGICRWKDSVVEETLKRKLP----- 170

QY 156 KXLWPRCLKXQXTLYELHYXXITFGKXXFCYXXPNCNACPM 198

Db 171 KELWS-----DAHHYIFFGR-YHCKARNPECPTCPL 201

Search completed: November 8, 2002, 19:26:43

Job time : 17.0327 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

[illegible]

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	166.5	60.5	1309	10	Q9SR66	Q9sr66 arabidopsis
2	163.5	59.5	1207	10	Q9S706	Q9sj96 arabidopsis
3	162	58.9	917	10	Q04948	Q9498 arabidopsis
4	49.5	18.0	1017	10	Q9Z669	Q91z69 arabidopsis
5	47	17.1	541	5	Q09617	Q09617 caenorhabdi
6	46	16.7	207	11	Q9D223	Q9d223 mus musculu
7	46	16.7	921	4	Q9Y4C2	Q9y4c2 homo sapien
8	45	16.4	1131	16	Q9PPT3	Q9ppt3 ureaplasma
9	44.5	16.2	373	10	Q9PNQ3	Q9fng3 arabidopsis
10	44	16.0	1141	3	Q9P3U0	Q9p3u0 schizosacch
11	43.5	15.8	327	10	Q9SZ60	Q9sz60 arabidopsis
12	43.5	15.8	788	10	Q9C9V5	Q9c9v5 arabidopsis
13	43.5	15.8	765	10	Q93YU0	Q93yu0 arabidopsis
14	43	15.6	120	17	Q8ZW64	Q8zw64 pyrobaculum
15	43	15.6	128	10	Q8RWM5	Q8rwm5 arabidopsis
16	43	15.6	217	10	Q9C9E0	Q9c9e0 arabidopsis

```

117 VDTNNGRIYVRXGVPLXPLPXXXXXHXIXLXPPXXXXXOKXLPRLCKLXQTXLYELHYX 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
625 VDTNNGRIANVLGLPPLPPLPVGQVQHQLFE-----YELHYQ 661
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
177 XITEGKXXFCTKXXXPNCNACPMXXXXXECXXXXXA 209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
662 MIIFGK-VFTKTIPIPCNACPM-KSECKYFASA 692
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

DE Hypothetical 82.2 kDa protein.
GN T23K23.26.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buchler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC012563; AAGS2018.1; -.
DR HSSP; P12931; IFMK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 738 AA; 82220 MW; B2B65569A52D6145 CRC64;

Query Match 15.8%; Score 43.5; DB 10; Length 738;
Best Local Similarity 45.8%; Pred. No. 75;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Qy 58 KGSVVDSVXGVFLTONDXXSXA 81
:|:::| | | | | | | | | | |
Db 333 RGNLIQSPFGVFLC-NDDKSSKA 355

RESULT 13
Q93YU0
ID Q93YU0 PRELIMINARY; PRT; 765 AA.
AC Q93YU0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 85.1 kDa protein.
GN ATIG67890.
OS Arabidopsis thaliana (Mouse-ear cress).

Query Match 15.8%; Score 43.5; DB 10; Length 738;
Best Local Similarity 45.8%; Pred. No. 75;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Qy 58 KGSVVDSVXGVFLTONDXXSXA 81
:|:::| | | | | | | | | | |
Db 333 RGNLIQSPFGVFLC-NDDKSSKA 355

RESULT 13
Q93YU0
ID Q93YU0 PRELIMINARY; PRT; 765 AA.
AC Q93YU0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 85.1 kDa protein.
GN ATIG67890.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Bann J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene Atlg67890 (GI:15220577).";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059769; AAL24117.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 1.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00086; PAC; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN.1.
KW ATP-binding; Hypothetical protein; Transferase.
SQ SEQUENCE 765 AA; 85149 MW; 6651738FB4E347D4 CRC64;

Query Match 15.8%; Score 43.5; DB 10; Length 765;
Best Local Similarity 45.8%; Pred. No. 78;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Qy 58 KGSVVDSVXGVFLTONDXXSXA 81
:|:::| | | | | | | | | | |
Db 333 RGNLIQSPFGVFLC-NDDKSSKA 355

RESULT 14
Q8ZW64
ID Q8ZW64 PRELIMINARY; PRT; 120 AA.
AC Q8ZW64;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PAE1954.
GN PAE1954.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009849; AAL63838.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 13398 MW; 34DF06D062B6AD5E CRC64;

Query Match 15.6%; Score 43; DB 17; Length 120;
Best Local Similarity 38.9%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 54 FXWXKGSVVDSVXGVFLT 71
| | | | | | | | | | |

```



```
Q09617
ID Q09617 PRELIMINARY; PRT; 541 AA.
AC Q09617;
DT 01-JAN-1999 (TrEMBLrel. 09, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 61.0 kDa protein ZK892.3 in chromosome II.
GN ZK892.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Lloyd C.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z48638; CAA88567.1; -.
DR WormPep; ZK892.3; CE01726.
DR InterPro; IPR001589; Actbind_actin.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr.1.
DR PROSITE; PS00019; ACTININ.1; UNKNOWN_1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 541 AA; 61045 MW; A7A1D7EDB140AC6F CRC64;

Query Match 17.1%; Score 47; DB 5; Length 541;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 11; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

Qy 52 RFXXXKSGVDSVXGVFLTNQXDXSXAXXAXFP 90
| | | | | | | | | | | | | | | | | |
Db 194 RTFSNANARILVVFVATITRWDFASVCAISALIFP 232

RESULT 6
Q9D223
ID Q9D223 PRELIMINARY; PRT; 207 AA.
AC Q9D223;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE A230020K05RIK protein.
GN A230020K05RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HYPOTHALAMUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., StaUBLI F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RA Nature 409:685-690(2001).
```

```
DR EMBL: AK020704; BAB32184.1; -.
DR MGD: MGI:1924824; A230020K05RIK.
SQ SEQUENCE 207 AA; 22411 MW; 558A70EC798BB4C9 CRC64;

Query Match 16.7%; Score 46; DB 11; Length 207;
Best Local Similarity 40.9%; Pred. No. 6.7;
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 52 RFXXXKSGVDSVXGVFLTNQ 73
| | | | | | | | | | | | | | | | | |
Db 165 RVLFAFPGNLTVTSVAGVFTDN 186

RESULT 7
Q9Y4C2
ID Q9Y4C2 PRELIMINARY; PRT; 921 AA.
AC Q9Y4C2; O9BW63;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA0738 protein.
GN KIAA0738.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [2]
RP SEQUENCE OF 1-919 FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB018281; BAA34458.1; -.
DR EMBL: BC000609; AAH00609.1; -.
SQ SEQUENCE 921 AA; 102139 MW; 03D39FDF90561F74 CRC64;

Query Match 16.7%; Score 46; DB 4; Length 921;
Best Local Similarity 33.3%; Pred. No. 35;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 52 RFXXXKSGVDSVXGVFLTNQXDXXS 78
| | | | | | | | | | | | | | | | | |
Db 165 RVLFTFPGNLTVTSVAGVFTDNKGDT 191

RESULT 8
Q9PPT3
ID Q9PPT3 PRELIMINARY; PRT; 1131 AA.
AC Q9PPT3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical membrane lipoprotein.
GN UU557.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
```


[illegible]

1	166.5	60.5	1309	10	Q9SR66	Q9SR66 arabidops1s
2	163.5	59.5	1207	10	Q9SJQ6	Q9SJQ6 arabidops1s
3	162	58.9	917	10	Q49498	Q49498 arabidops1s
4	49.5	18.0	1017	10	Q9LZ69	Q9LZ69 arabidops1s
5	47	17.1	541	5	Q09617	Q09617 caenorhabd1
6	46	16.7	207	11	Q9D223	Q9D223 mus musculu
7	46	16.7	921	4	Q9Y4C2	Q9Y4C2 homo sapien
8	45	16.4	1131	16	Q9PPT3	Q9PPT3 ureaplasma
9	44.5	16.2	373	10	Q9ENQ3	Q9ENQ3 arabidops1s
10	44	16.0	1141	3	Q9P300	Q9P300 schizosacch
11	43.5	15.8	327	10	Q9SZ60	Q9SZ60 arabidops1s
12	43.5	15.8	738	10	Q9C9V5	Q9C9V5 arabidops1s
13	43.5	15.8	765	10	Q93YU0	Q93YU0 arabidops1s
14	43	15.6	120	17	Q8Z464	Q8Z464 pyrobaculum
15	43	15.6	128	10	Q8RMW5	Q8RMW5 arabidops1s
16	43	15.6	217	10	Q9C9E0	Q9C9E0 arabidops1s

.....

Qy 53 XFXXWKGSVVD SVXGVFLTQN-XDXXSSXAXMXXAXXFP 90

THIS PAGE BLANK (USPTO)

Db 118 LIKLPGRKCANIVLAYGFGIPAPVDTHVYRISRRLGLAPWDASPEVEERL----- 171

Qy 151 XXXXOKXLWRLCKLXQXTLYELHYXXITFGKXXFCTKXXPNACPM 198

Db 172 -----KSLIPR-----EEWIY-VNHAMVDHGK-SVCKPIKPRCWEPL 207

Search completed: November 8, 2002, 19:28:38
Job time : 32.5621 secs

```

DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Endonuclease III.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales;
OC Methanobacteriacea: Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Alldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakhar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000855; AAB85267.1; -
DR HSP: P20625; 2ABK
DR InterPro: IPR004036; EndoIII_HhH.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR003583; HHH_1.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00525; FES; 1.
DR SMART: SM00278; HhH; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 233 AA; 27012 MW; 10F6163AAE98395B CRC64;

Query Match 12.0%; Score 83.5; DB 17; Length 233;
Best Local Similarity 26.9%; Pred. No. 1.4e-05;
Matches 29; Conservative 10; Mismatches 50; Indels 19; Gaps 4;

QY 91 LLKXXGXGLKSECVRLLLXXXXAFVDTNVGRXVVRGXVPLPLPXXQXHXLLXXYPX 150
DB 127 LKLFQVGRKTNCLVYVAFGRPALPVDTHVRIENRGLVD-TRTPEERALMKVIP- 184
QY 151 XXXXQXKLWPLRLCKLXQXTLYELHYXXITFGKXXFCCTKXXPCNACPM 198
DB 185 -----REYW-----IELNDLWQFGQ-DICRPLGPRHECPI 215

RESULT 14
Q971F2
ID Q971F2 PRELIMINARY; PRT; 224 AA.
AC Q971F2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2002 (TReMBLrel. 20, Last annotation update)
DE Putative endonuclease III.
GN ST1401.
OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic

```

```

RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000986; BAB66468.1; -
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003583; HHH_1.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00278; HhH; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; UNKNOWN_1.
KW Endonuclease; Hypothetical protein; Complete proteome.
SQ SEQUENCE 224 AA; 25629 MW; E2BE3A0762135550 CRC64;

Query Match 12.0%; Score 83; DB 17; Length 224;
Best Local Similarity 25.7%; Pred. No. 1.6e-05;
Matches 27; Conservative 13; Mismatches 47; Indels 18; Gaps 3;

QY 92 LXXGXGLKSECVRLLLXXXXAFVDTNVGRXVVRGXVPLPLPXXQXHXLLXXYPX 151
DB 123 LTVDIGEKATADVIVNCFKCKFFVDPVTHIKRVMSRLGILGSKP-----QYKEIADF--- 174
QY 152 XXXQXKLWPLRLCKLXQXTLYELHYXXITFGKXXFCCTKXXPCNAC 196
DB 175 -----FTSSLNEDELLELHQLLHGRKT-CTAKKPLCDKC 209

RESULT 15
Q59167
ID Q59167 PRELIMINARY; PRT; 222 AA.
AC Q59167;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 222AA long hypothetical endonuclease III.
GN PH1498.
OS Pyrococcus horikoshii.
OC Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000006; BAA30606.1; -
DR HSP: P20625; 2ABK.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR004036; EndoIII_HhH.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR003583; HHH_1.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00525; FES; 1.
DR SMART: SM00278; HhH; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; UNKNOWN_1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 222 AA; 25553 MW; DD096F91C969DF0C CRC64;

Query Match 11.8%; Score 82; DB 17; Length 222;
Best Local Similarity 26.9%; Pred. No. 2.4e-05;
Matches 29; Conservative 10; Mismatches 51; Indels 18; Gaps 4;

QY 91 LLKXXGXGLKSECVRLLLXXXXAFVDTNVGRXVVRGXVPLPLPXXQXHXLLXXYPX 150

```

Db. 162 EYDELSRW-FMERLPKDYLFHLKLIQFGR-DVCRARNPKGQCPI-GAKCPSFKSA 217

RESULT 10

Q9A623 PRELIMINARY; PRT; 241 AA.

AC Q9A623; 17, Created

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Endonuclease III family protein.

GN CC2272.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 19089 / CB15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub M.R.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005897; AAK24243.1; -

DR TIGR; CC2272; -

DR InterPro; IPR003265; Endo_3c.

DR InterPro; IPR003651; Fes_bind.

DR Pfam; PF00730; HHH-GPD; 1.

DR SMART; SM00478; ENDO3c; 1.

DR SMART; SM00525; FES; 1.

KW Complete proteome.

SQ SEQUENCE 241 AA; 26421 MW; EB43FBA976736F1C CRC64;

Query Match 12.5%; Score 86.5; DB 16; Length 241;

Best Local Similarity 24.2%; Pred. NO. 4.5e-06;

Matches 31; Conservative 9; Mismatches 73; Indels 15; Gaps 3;

Qy 72 GXIDLEWRXXXXXXLLXXXXGLKSECV-RLXLXXXAFVDTNVGRIVRXGX 130

Db 106 GWLSLSHLKLEVDQARWELQALPGVGKVAACVLFNSDLAMRALVVDTHVDRVARRIGL 165

Qy 131 VPLXPLPXXXXXHXLLXXPPXXXXXKXLPRLKXQXTLVHLYXXITFGKXXFCTKXX 190

Db 166 V-----GSGDNTTNTHTLMAMAPASW-----TADDLFELHLMKRLGQMLCGAEA 211

Qy 191 PNCNACPM 198

Db 212 PKCGACP 219

RESULT 11

Q9V0W6 PRELIMINARY; PRT; 222 AA.

AC Q9V0W6; 13, Created

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Endonuclease III (NTH).

GN NTH OR PAB0459.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ORSAY;

RA Heilig R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ248285; CAB49586.1; -

DR HSSP; P20625; 2ABK.

DR InterPro; IPR004035; EndoIII_FCL.

DR InterPro; IPR004036; EndoIII_HhH.

DR InterPro; IPR003265; Endo_3c.

DR InterPro; IPR003651; Fes_bind.

DR InterPro; IPR003583; HHH_1.

DR Pfam; PF00730; HHH-GPD; 1.

DR SMART; SM00478; ENDO3c; 1.

DR SMART; SM00525; FES; 1.

DR SMART; SM00278; HhH; 1.

DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.

DR PROSITE; PS01155; ENDONUCLEASE_III_2; UNKNOWN_1.

KW Endonuclease; Complete proteome.

SQ SEQUENCE 222 AA; 25541 MW; 27F0B5FC83C0DC6 CRC64;

Query Match 12.4%; Score 86; DB 17; Length 222;

Best Local Similarity 26.9%; Pred. NO. 5.1e-06;

Matches 29; Conservative 11; Mismatches 50; Indels 18; Gaps 4;

Qy 91 LLXXXGXLKSECVRLXLXXXAFVDTNVGRIVRXGXVPLXPLPXXXXXHXLLXXYPX 150

Db 118 LKLPFGIGRKCANIVLAYGFGIPAVPVDTHVYRRLGLAPWDASPEVEERL----- 171

Qy 151 XXXXOKLWPLRLKXQXTLVHLYXXITFGKXXFCTKXXPNCNACPM 198

Db 172 -----KELIPR-----EEWIY-VNHAMVDHGK-SVCRPIKPRCDECP 207

RESULT 12

Q8UIH8 PRELIMINARY; PRT; 225 AA.

ID Q8UIH8;

AC Q8UIH8;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Glycosylase putative, mutY-nth family.

GN PF1229.

OS Pyrococcus furiosus.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI_TaxID=2261;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VCI / DSM 3638 / ATCC 43587 / JCM 8422;

RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;

RT "The complete sequence of the Pyrococcus furiosus genome.";

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE010230; AAL81353.1; -

KW Complete proteome.

SQ SEQUENCE 225 AA; 26157 MW; DA48762AC33BE4AE CRC64;

Query Match 12.1%; Score 84; DB 17; Length 225;

Best Local Similarity 25.9%; Pred. NO. 1.1e-05;

Matches 28; Conservative 11; Mismatches 51; Indels 18; Gaps 4;

Qy 91 LLXXXGXLKSECVRLXLXXXAFVDTNVGRIVRXGXVPLXPLPXXXXXHXLLXXYPX 150

Db 127 LMKLPFGIGRKCANIVLAYGFGKPAIPVDTHVYRRLGLAPINSTPEKVE----- 177

Qy 151 XXXXOKLWPLRLKXQXTLVHLYXXITFGKXXFCTKXXPNCNACPM 198

Db 178 --EILKTLIP-----VEEWIY-VNHAMVDHGK-SICRPIKPKCELCP 216

RESULT 13

O26858 PRELIMINARY; PRT; 233 AA.

ID O26858

AC O26858;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

[illegible]

```
RESULT 5
Q9SJO6
ID Q9SJO6 PRELIMINARY; PRT; 1207 AA.
AC Q9SJO6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AT2G36490 protein.
GN AT2G36490.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006919; AAD24633.1; -.
DR InterPro; IPR003651; Fes_bind.
DR SMART; SM00525; FES; 1.
SQ SEQUENCE 1207 AA; 135916 MW; AADB6C9EE94DDF67 CRC64;

Query Match 20.1%; Score 139; DB 10; Length 1207;
Best Local Similarity 25.1%; Pred. No. 3.4e-14;
Matches 50; Conservative 0; Mismatches 65; Indels 84; Gaps 5;

QY 11 EXXXXXXXXXXXXXXXXXXXXXXXXXXGXGKXECVRLXXXXXXXXXXXXXFLXXXXXX 70
Db 846 EXTRSTMDVDMKAIFAADVKEVAETIKSRGMNHKLAERIQVLTLMNKTMOGFLDLVND 905
QY 71 GXIDLEWLXXXXXXXXXXXXXXXXXXXXXGXGKXECVRLXXXXXXXXXXXXXFLXXXXXX 130
Db 906 HGSIDLEWL-----D 916
QY 131 VPLXPLPXXXXXHLXYPXXXXXKXLMPLRCKLXQXTLYELHYXXITFGKXXCTKXX 190
Db 917 VP-----PDKAK-----YELHYQMIFGK-VFCTKSK 942
QY 191 PNCNACPMXXECXXXXXA 209
Db 943 PNCNACPM-KGECRHFA 960

RESULT 6
Q9RR00
ID Q9RR00 PRELIMINARY; PRT; 259 AA.
AC Q9RR00;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Endonuclease III.
GN DR2438.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.

Query Match 16.2%; Score 112.5; DB 16; Length 259;
Best Local Similarity 25.2%; Pred. No. 2.1e-10;
Matches 35; Conservative 10; Mismatches 49; Indels 45; Gaps 3;

OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE002073; AAF11977.1; -.
DR TIGR; DR2438; -.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; Fes_bind.
DR Pfam; PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 259 AA; 28946 MW; AA07354719093547 CRC64;

Query Match 16.2%; Score 112.5; DB 16; Length 259;
Best Local Similarity 25.2%; Pred. No. 2.1e-10;
Matches 35; Conservative 10; Mismatches 49; Indels 45; Gaps 3;

QY 72 GXIDLEWLXXXXXXXXXXXXXXXXXXXXXGXGKXECVRLXXXXXXXXXXXXXFLXXXXXX 131
Db 120 GGYDLDFLEDLPVKDALKWLTDLPGVGKTSALVLLFNYPVPVDTTHRVSTRVGI 179
QY 132 P-----LXPLPXXXXXHLXYPXXXXXKXLMPLRCKLXQXTLYELHYXX 177
Db 180 PRMGEQAHRALLALLPPDPY-----LYELHINF 209
QY 178 ITFGKXXCTKXXPCNAC 196
Db 210 LSHGR-QVCTWTRPKGKC 227

RESULT 7
P95945
ID P95945 PRELIMINARY; PRT; 236 AA.
AC P95945;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Endonuclease III (DNA endonuclease III, probable) (NTH-1)
DE (EC 4.2.99.18).
GN NTH-1 OR SSO0116 OR C04006.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
RA Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA Doolittle W.F., Ragan M.A., Charlebois R.L.;
RT "Organizational characteristics and information content of an archaeal
genome: 156 kb of sequence from Sulfolobus solfataricus P2.";
RL Mol. Microbiol. 22:175-191(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 28.5621 Seconds
(without alignments)
1659.222 Million cell updates/sec

Title: US-09-840-743-72

Perfect score: 693

Sequence: 1 WXXXXXXXEXXXXXXX.....XXXXXXXXXXXXXXXXXXXX 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTRMBL21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	335.5	48.4	1309	10	Q9SR66
2	228	32.9	555	10	Q9L268
3	202	29.1	119	10	Q94LX6
4	201.5	29.1	917	10	Q49498
5	139	20.1	1207	10	Q9SUQ6
6	112.5	16.2	259	16	Q9RRQ0
7	89.5	12.9	236	17	P95945
8	89.5	12.9	265	17	Q9HS56
9	87.5	12.6	223	17	Q9P9L7
10	86.5	12.5	241	17	Q9A623
11	86	12.4	222	17	Q9V0W6
12	84	12.1	225	17	Q8UIH8
13	83.5	12.0	233	17	O26858
14	83	12.0	224	17	Q971F2
15	82	11.8	222	17	O59167
16	78.5	11.3	338	16	Q9RVU4

17	76	11.0	232	16	Q9WVL3
18	74.5	10.8	572	3	Q8R957
19	74.5	10.8	572	3	Q9HF16
20	74	10.7	218	16	Q9CGM5
21	72.5	10.5	216	17	Q8THA8
22	72	10.4	219	16	Q92AB4
23	72	10.4	219	16	Q8Y608
24	70.5	10.2	233	17	Q8TY08
25	70.5	10.2	263	16	Q92T20
26	70	10.1	214	16	Q9ZLP6
27	70	10.1	218	16	O25307
28	69.5	10.0	213	16	O66636
29	69	10.0	218	16	Q9KC76
30	68.5	9.9	264	17	Q9YFV0
31	68	9.8	219	16	Q99U37
32	68	9.8	220	2	Q9EUT1
33	67.5	9.7	282	17	Q8TK57
34	67	9.7	209	17	O28581
35	66.5	9.6	206	17	Q8TVM6
36	64.5	9.3	215	17	Q97AJ2
37	64	9.2	224	2	Q9V1H0
38	63	9.1	308	5	Q9V1H0
39	62.5	9.0	225	16	Q9RXM1
40	62	8.9	297	16	Q9CB70
41	61.5	8.9	292	16	Q97FM4
42	61	8.8	209	16	Q97OE0
43	61	8.8	1072	5	Q9W592
44	60.5	8.7	211	16	Q97L73
45	60.5	8.7	238	16	Q98GU1

ALIGNMENTS

RESULT 1

Q9SR66	PRELIMINARY;	PRT;	1309 AA.
ID	Q9SR66		
AC	Q9SR66;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	T22K18.18 protein.		
GN	T22K18.18.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,		
RA	Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,		
RA	Bowman C.L., White O., Nierman W.C., Fraser C.M.;		
RT	"Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";		
RL	Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AC010927; AAF04422.1;		
DR	InterPro; IPR003265; Endo.3c.		
DR	InterPro; IPR003651; Fes_bind.		
DR	Pfam; PF00730; HH-GPD; I.		
DR	SMART; SM00478; ENDO3c; I.		
DR	SMART; SM00525; FES; I.		
SQ	SEQUENCE 1309 AA; 148451 MW; 3B21E4439BA4G73 CRC64;		

Query Match	48.4%;	Score 335.5;	DB 10;	Length 1309;
Best Local Similarity	47.4%;	Pred. No. 3.1e-47;		
Matches	82;	Conservative	0;	Mismatches 56;
Indels	35;	Gaps	4;	
Qy	37	IXXRGXXXXXXRIXXXXXXXXXXXXXIDLEWLXXXXXXXXXXXXXXXXXXG 96		
Db	832	IIKRGMMNLAERI-----KAFNLRLVKKHGSDLEWLDPDPDKAKAYLLSING 881		
Qy	97	XGLSXECVRLLLXXXXXAFVDNTNVGRIVRXGVPLPLPXXXXXHXLLXPXXXXXQK 156		

C:Superfamily: hypothetical protein HI0107

Query Match 15.6%; Score 43; DB 2; Length 434;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 62 VDSVXGVFLTONXD 75
|::| | | | | | | |
DB 386 VDTIGGWFLTONYE 399

RESULT 12

Al1723

hypothetical protein homolog lin2334 [imported] - *Listeria innocua* (strain Cl111262)

C:Species: *Listeria innocua*

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002

C:Accession: Al1723

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluteter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: Al1723

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <GLA>

A:Cross-references: GB:AL592022; PIDN:CA97561.1; PID:gl6414857; GSPDB:GN00178

A:Experimental source: strain Cl111262

C:Genetics:

A:Gene: lin2334

C:Superfamily: hypothetical protein HI0107

Query Match 15.6%; Score 43; DB 2; Length 434;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 62 VDSVXGVFLTONXD 75
|::| | | | | | | |
DB 386 VDTIGGWFLTONYE 399

RESULT 13

T29718

hypothetical protein K11G9.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000

C:Accession: T29718

R:Greco, T.; Bradshaw, H.

A:Description: The sequence of *C. elegans* cosmid K11G9.

A:Reference number: 220672

A:Accession: T29718

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-565 <GR>

A:Cross-references: EMBL:U64853; PIDN:AAB04976.1; GSPDB:GN00023; CESP:K11G9.1

A:Experimental source: strain Bristol N2; clone K11G9

C:Genetics:

A:Gene: CESP:K11G9.1

A:Map position: 5

A:Introns: 31/3; 114/3; 144/2; 282/1; 352/1; 460/1

C:Superfamily: cholinesterase; cholinesterase homology

Query Match 15.6%; Score 43; DB 2; Length 565;
Best Local Similarity 30.8%; Pred. No. 34;
Matches 12; Conservative 4; Mismatches 17; Indels 6; Gaps 1;

QY 57 WKGSVVDVSGV-----FLTONDXSSXAXMXXYF 89
|::| | | | | | | |
DB 374 YKEDVDDVEGIRKEIFETTKNISEGDEEAIMRKAEEF 412

RESULT 14

H72278

alpha-mannosidase-related protein - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: H72278

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: H72278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-832 <ARN>

A:Cross-references: GB:AE001779; GB:AE000512; NID:g4981777; PIDN:AAD36306.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1231

Query Match 15.6%; Score 43; DB 2; Length 832;

Best Local Similarity 64.3%; Pred. No. 52;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 57 WKGSVVDVSGVFL 70

||||| -| |||||

DB 160 WKSGDGTAVQGVFL 173

RESULT 15

AE2799

HLVD family secretion protein [imported] - *Agrobacterium tumefaciens* (strain C58, Dup

C:Species: *Agrobacterium tumefaciens*

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AE2799

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AE2799

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-432 <KUR>

A:Cross-references: GB:AE008686; PIDN:AAL42811.1; PID:gl7740257; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atul815

A:Map position: circular chromosome

Query Match 15.3%; Score 42; DB 2; Length 432;

Best Local Similarity 31.0%; Pred. No. 38;

Matches 9; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 58 KGSVVDVSGVFLTONDXSSXAXMXXYA 86

|::| | | | | | | |

DB 212 KAKIVSPIDGVILTRVDPGATVASSLNA 240

Search completed: November 8, 2002, 19:26:41

Job time : 8.88235 secs

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick,A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C96748
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <STO>
A:Cross-references: GB:AE005173; NID:g6730759; PIDN:AAF27148.1; GSPDB:GN00141
C:Genetics:
A:Gene: T10D10.11
A:Map position: 1

	Query Match	15.6%	Score 43;	DB 2:	Length 217;
	Best Local Similarity	26.1%;	Pred. No. 12;		
Matches	12;	Conservative	5;	Mismatches	21;
				Indels	8;
				Gaps	1;

QY 52 RXFXXWKGSVVDS-----VXGVFLTONXXXXSSAXMXXXXFF 89
I IIIII:I I : I : I
Db 159 RYLPTWKGNVIDAEMENPGHVVGMSLSVNAQGSGFIGAKSGAGDF 204

RESULT 10
G83962
hypothetical protein BH2503 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83962
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: G83962
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BAB06222.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2503
C:Superfamily: conserved hypothetical protein HI1714

	Query Match	15.6%	Score 43;	DB 2:	Length 294;
	Best Local Similarity	38.9%;	Pred. No. 17;		
Matches	7;	Conservative	5;	Mismatches	6;
				Indels	0;
				Gaps	0;

QY 58 KGVSVDVXGVFLTONXD 75
I::I I I I I I
Db 3 KGTIVKALSGFYVQNED 20

RESULT 11
AH1353
hypothetical protein lmo2232 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
C:Accession: AH1353
R:Glaser, P.; Frangul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fslhi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1353
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-434 <GLA>
A:Cross-references: GB:NC_003210; PTDN:CAD000310.1; PID:gl6411702; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2232

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	163.5	59.5	1207	2	D84781		hypothetical prote
2	162	58.9	917	2	T05430		hypothetical prote
3	49.5	18.0	1017	2	T48452		hypothetical prote
4	47	17.1	442	2	E69958		conserved hypothet
5	47	17.1	541	2	T28069		hypothetical prote
6	45	16.4	1131	2	F82875		hypothetical prote
7	43.5	15.8	327	2	T06602		hypothetical prote
8	43.5	15.8	738	2	F96701		hypothetical prote
9	43	15.6	217	2	C96748		unknown protein T1
10	43	15.6	294	2	G83962		hypothetical prote
11	43	15.6	434	2	AH1353		hypothetical prote
12	43	15.6	434	2	AI1723		hypothetical prote
13	43	15.6	565	2	T29718		hypothetical prote
14	43	15.6	832	2	H72278		alpha-mannosidase-
15	42	15.3	432	2	AE2799		HLyd family secret
16	42	15.3	437	2	G97578		probable periplasm
17	42	15.3	868	2	G82193		aminopeptidase N V
18	42	15.3	894	2	T01385		probable phosphor
19	41.5	15.1	723	2	T30094		hypothetical prote
20	41	14.9	108	1	WRBP69		early protein gp13
21	41	14.9	108	1	WRBP13		early protein gp13
22	41	14.9	108	2	JN0031		early protein gp13
23	41	14.9	252	2	D81052		acylttransferase, p
24	41	14.9	253	2	G69199		endonuclease III-r
25	41	14.9	260	2	H69874		conserved hypothet
26	41	14.9	280	2	D84968		hypothetical prote
27	41	14.9	319	2	T35933		hypothetical prote
28	41	14.9	444	2	F69826		hemolysin homolog
29	41	14.9	638	2	T24676		hypothetical prote

THIS PAGE BLANK (USPTO)

```

RESULT_15
ABB48015
ID   ABB48015 standard; Protein; 219 AA.
XX
AC   ABB48015;
XX
DT   05-FEB-2002 (first entry)
XX
DE   Listeria monocytogenes protein #719.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
    vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 200LWO-FR01118.
XX
PR   11-APR-2000; 200OFR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI   Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI   Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI   Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI   Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI   Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI   Maqueda E, De Pablo B, Wehnald K, Kaerst U, Entian K, Hauf J;
PI   Rose M, Voss H;
XX
WPI: 2002-010914/01.
XX
Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
and prevention of Listeria and related bacterial infections, and
related polypeptides -
XX
Claim 6; SEQ ID No 720; 192pp; French.
XX
The present invention relates to the genome sequence of Listeria
monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
it are useful for selecting probes and primers for detecting genes in L.
monocytogenes and related organisms, and for studying genetic
polymorphisms and other genomes. The present sequence is a protein
expressed from the genome sequence of the present invention. Proteins
antibodies, identification of L. monocytogenes and related organisms, and
for biosynthesis and biodegradation, especially biosynthesis of Vitamin
B12. The genome sequence and proteins encoded by it are also useful for
selecting compounds that regulate gene expression and cell replication
and modulate L. monocytogenes-related diseases. In addition, the genome
sequence and proteins encoded by it are useful in pharmaceutical and
vaccines compositions for the treatment or prevention of infections by L.
monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence      219 AA;
Query Match          10.4%; Score 72; DB 23; Length 219;
Best Local Similarity 25.2%; Pred. No. 0.0017;
Matches 26; Conservative 8; Mismatches 51; Indels 18; Gaps 3;
QY 96 GXGLKSECEVRLLLXXXAFPDNVGRGXVRXGVNPLXPFPXXXQHXLXXYPXXXXXQ 155
Db 117 GVGRKTANVLVSFGVGVPATIAVDHVERISKRLGICRWKDSVVEETLKRLP----- 170
QY 156 KXLWPRECLKXQXTLYELHYXXITTFGRXXXTFKTKXXPCNCACP 198

```


CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents a DMT-like protein.
XX
SQ Sequence 145 AA;
Query Match 20.3%; Score 141; DB 23; Length 145;
Best Local Similarity 45.1%; Pred. No. 2.8e-15;
Matches 37; Conservative 0; Mismatches 35; Indels 10; Gaps 1;
QY 37 IXRGXXXXLXXRXXXXXXXFLXXXXXXGIDLEWLRXXXXXXXKXLLXXG 96
DB 74 IRRGMNVLSRI-----QEFNLVRDHGSIDLEWLRDIPPSAKDVLISRG 123
QY 97 XGLKSECVRLLLXXAAPPVD 118
DB 124 LGLKSECVRLLLTLHLAAPPVD 145
RESULT 11
AAU72747
ID AAU72747 standard; Protein; 276 AA.
XX
AC AAU72747;
XX
DT 26-FEB-2002 (first entry)
XX
DE Corn DMT3 protein.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Zea mays.
XX
PN WO200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
DR WPI; 2002-055307/07.
XX
PT New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
XX
PS Disclosure; Page 86; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of

CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents a DMT-like protein.
XX
SQ Sequence 276 AA;
Query Match 15.3%; Score 106; DB 23; Length 276;
Best Local Similarity 27.3%; Pred. No. 4.2e-09;
Matches 38; Conservative 12; Mismatches 65; Indels 24; Gaps 3;
QY 72 GXIDLEWLRXXXXXXXGLKXGKXECVRLLLXXAAPPVDNVRGIRVVRGXV 131
DB 147 GKICLEYLRELSVDVKELSRFKGIGPKTVACVLMFYLOKDDDFVDTLRLTKAMGW 206
QY 132 PLXPLPXXXXXHXLLXXYPXXXXXOKXILMPRLCKLXQTLVELHYHXXITFGKX-XFCT--- 187
DB 207 PATASREKAYIHLNN-----KIPDDLKFDLNCFLFVTHGKLCQSCTKKV 249
QY 188 ---KXXPCNACPMXXYEC 203
DB 250 GSKRKSNSACPLAGYCC 268
RESULT 12
AAU72746
ID AAU72746 standard; Protein; 166 AA.
XX
AC AAU72746;
XX
DT 26-FEB-2002 (first entry)
XX
DE Corn DMT2 protein.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Zea mays.
XX
PN WO200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
DR WPI; 2002-055307/07.
XX
PT N-PSDB; AAS96697.
XX
PS New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
XX
PS Disclosure; Page 85; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is

CC transcription of target genes by demethylation. The present
sequence represents a DMT-like protein.

XX

SQ Sequence 179 AA;

Query Match 36.8%; Score 255; DB 23; Length 179;
Best Local Similarity 63.0%; Pred. No. 2.5e-34;
Matches 51; Conservative 0; Mismatches 28; Indels 2; Gaps 2;

QY 129 GXVPLPPLPXXXQHXLXYPXXXXKXLPRLCKLQXTLYELHXXITFGKXXFCTK 188
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 GWPLQLPSLQHLLLELPVLESIOKYLPRLCKLDQETLYELHQMITFGK-VFCTK 60
| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 189 XPNACNCPMXXECXXXXSA 209
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 SKPNCNCPM-RAEGRHFASA 80
| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
AAU72749
ID AAU72749 standard; Protein; 145 AA.
XX
XX AC AAU72749;
AC
XX
DT DT 26-FEB-2002 (first entry)
XX
XX Wheat DMT2 protein.
DE
KW Demeter; DMT: Atropos; ATR: 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Triticum sp.
XX
PN WO200180626-A1.
PP
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
PR
XX 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, TatariNova TV;
XX
XX WPI: 2002-055307/07.
DR N-PSDB; AAS96699.
PT
PT New polynucleotide that control plant development comprising a sequence
PS having a specific homology to DEMETER domains A,B or C -
PS Disclosure; Page 87; 109pp; English.
XX
XX The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B,
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the

[illegible]

Db 1554 YLWPRLCKLDQRTLYELHYQMITEGK - VFCTKSKPNCNACPM - RAECKHFASA 1604

RESULT 5

AAU72752
ID AAU72752 standard; Protein; 298 AA.

XX
AC AAU72752;

DT 26-FEB-2002 (first entry)

DE Soybean DMT1 protein.

Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.

Glycine max.

XX
PN WO200180626-A1.

XX PD 01-NOV-2001.

XX
PF 23-APR-2001; 2001WO-US13059.

XX
PR 21-APR-2000; 2000US-0553690.XX
PA - (REGC) UNIV CALIFORNIA.

PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;

DR WPI; 2002-055307/07.

DR N-PSDB; AAS96702.

New polynucleotide that control plant development comprising a sequence PT having a specific homology to DEMETER domains A,B or C -

XX PS Disclosure; Page 88; 109pp; English.

The invention relates to an isolated polynucleotide sequence or their complement encoding a polypeptide having a sequence at least 40% identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B or C or their combinations. Also included are an expression cassette comprising the polynucleotide or comprising a heterologous polynucleotide under the control of a promoter at least 70% identical to DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated region of DMT, a host cell comprising an exogenous polynucleotide encoding a DMT-like protein and a transgenic plant comprising a polynucleotide encoding a DMT-like protein. The expression cassette is useful for modulating transcription. The method comprises introducing the cassette into a host cell preferably *Agrobacterium* by sexual cross, and selecting a host cell with modulated transcription, where the protein is capable of exhibiting at least one of the following biological activities, which include enhanced expression of the protein in a plant results in a delay in flowering time, introduction of the protein into a cell results in modulation of methylation of chromosomal DNA in the cell, reduction of expression of the protein in a plant results in enhanced endosperm development and expressing of the protein in an *Arabidopsis* leaf results in expression of the MEDEA gene. The polynucleotide is useful for detecting a nucleic acid in a sample. DEMETER is related to 5-methylcytosine glycosylases and regulates transcription of target genes by demethylation. The present sequence represents a DMT-like protein.

AA	Sequence	298 AA;
SQ		

Query Match 62.3%; Score 432; DB 23; Length 298;
Best Local Similarity 56.6%; Pred. No. 8.4e-64;

37 IXRGXXXXLXXRXIXXXXXXXXXXXXXFLXXXXXGXIDLEWLRXXXXXXXXXXXXLXXGX 96

21 IKERGMNNRLADRI-----KNFNLRLVEEHGSDLEWL RDPDKAKYLLSIRG 70

[illegible][illegible]

XX PS Disclosure; Page 67; 109pp; English.

XX CC The invention relates to an isolated polynucleotide sequence or their

XX CC complement encoding a polypeptide having a sequence at least 40%

XX CC identical to DMT (DEMEter, previously known as ATROPOS (ATR)) Domain A, B

XX CC or C or their combinations. Also included are an expression

XX CC cassette comprising the polynucleotide or comprising a heterologous

XX CC polynucleotide under the control of a promoter at least 70% identical to

XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

XX CC region of DMT, a host cell comprising an exogenous polynucleotide

XX CC encoding a DMT-like protein and a transgenic plant comprising a

XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

XX CC useful for modulating transcription. The method comprises introducing

XX CC the cassette into a host cell preferably Agrobacterium by sexual

XX CC cross, and selecting a host cell with modulated transcription, where

XX CC the protein is capable of exhibiting at least one of the following

XX CC biological activities, which include enhanced expression of the

XX CC protein in a plant results in a delay in flowering time, introduction of

XX CC the protein into a cell results in modulation of methylation of

XX CC chromosomal DNA in the cell, reduction of expression of the protein in a

XX CC plant results in enhanced endosperm development and expression of the

XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.

XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.

XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates

XX CC transcription of target genes by demethylation. The present

XX CC sequence represents a DMT-like protein.

XX SQ Sequence 1413 AA;

Query Match 63.5%; Score 440; DB 23; Length 1413;

Best Local Similarity 49.7%; Pred. No. 2e-64;

Matches 99; Conservative 0; Mismatches 98; Indels 2; Gaps 2;

Qy 11 EXXXXXXXXXXXXXXXXXXXXGXGGLKXGKXECVRLLLXXXAFPPVDTNVGRXVXGX 70

Db 874 EXRSTMDVDMKAIKRAADVKEVETIKSRGMNHKLAEIRIQLTNMKMGFDLRLVND 933

Qy 71 XGIDLEWLRLXXXXXXLLXXXGXGGLKXGKXECVRLLLXXXAFPPVDTNVGRXVXGX 130

Db 934 HGSIDLEWLRLVPPDKAKLEYLLSFNGLGLKXGKXECVRLLLTHHLAFPPVDTNVGRXVXGX 993

Qy 131 VPLXPLPXXXXXHXLLXYPXXXXXOKXLPRLCKLXOXTLYELHYXXITFGKXXFXCTXXX 190

Db 994 VPLQPLPESLQHLLEWYPMLESIQKYLWPRCLKLDQKTLXELHYQMITFGK-VFCTRSK 1052

Qy 191 PNCNACPMXXECXXXXXA 209

Db 1053 PNCNACPM-KGECRHFASA 1070

RESULT 2

AAU72741

ID AAU72741 standard; Protein; 1114 AA.

XX AC AAU72741;

XX DT 26-FEB-2002 (first entry)

XX DE Arabidopsis DMT4 (IDMT4) protein.

XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;

XX KW DNA demethylation; transgenic plant; transcription modulation;

XX KW flowering time; endosperm development; MEDEA.

XX OS Arabidopsis Thaliana.

XX PN WO200180626-A1.

XX PD 01-NOV-2001.

XX PF 23-APR-2001; 2001WO-US13059.

XX

PR 21-APR-2000; 2000US-0553690.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;

XX DR WPI; 2002-055307/07.

XX DR N-PSDB; AAS96694.

XX PT New polynucleotide that control plant development comprising a sequence

XX PT having a specific homology to DEMETER domains A,B or C

XX PS Disclosure; Page 75; 109pp; English.

XX CC The invention relates to an isolated polynucleotide sequence or their

XX CC complement encoding a polypeptide having a sequence at least 40%

XX CC identical to DMT (DEMEter, previously known as ATROPOS (ATR)) Domain A, B

XX CC or C or their combinations. Also included are an expression

XX CC cassette comprising the polynucleotide or comprising a heterologous

XX CC polynucleotide under the control of a promoter at least 70% identical to

XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

XX CC region of DMT, a host cell comprising an exogenous polynucleotide

XX CC encoding a DMT-like protein and a transgenic plant comprising a

XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

XX CC useful for modulating transcription. The method comprises introducing

XX CC the cassette into a host cell preferably Agrobacterium by sexual

XX CC cross, and selecting a host cell with modulated transcription, where

XX CC the protein is capable of exhibiting at least one of the following

XX CC biological activities, which include enhanced expression of the

XX CC protein in a plant results in a delay in flowering time, introduction of

XX CC the protein into a cell results in modulation of methylation of

XX CC chromosomal DNA in the cell, reduction of expression of the protein in a

XX CC plant results in enhanced endosperm development and expression of the

XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.

XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.

XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates

XX CC transcription of target genes by demethylation. The present

XX CC sequence represents a DMT-like protein.

XX SQ Sequence 1114 AA;

Query Match 62.5%; Score 433; DB 23; Length 1114;

Best Local Similarity 55.5%; Pred. No. 2.3e-63;

Matches 96; Conservative 0; Mismatches 65; Indels 12; Gaps 3;

Qy 37 IXRGXXXXXXRX 96

Db 595 IKRGQFRLSERI-----LKFLNDEVNQNGNIDLEWLRNAPSHLVKRYLLETEG 644

Qy 97 XGLKXGKXECVRLLLXXXAFPPVDTNVGRXVXGXVPLXPLPXXXXXHXLLXYPXXXXXOK 156

Db 645 IGLKSAECVRLGLKHHAFPPVDTNVGRXVXGXVPLXPLPXXXXXHXLLXYPXXXXXOK 704

Qy 157 XLWPRCLKLXOXTLYELHYXXITFGKXXFXCTKXXPNACPMXXECXXXXXA 209

Db 705 YLWPRCLKLPQETLYELHYQMITFGK-VFCTKTIPNCNACPM-KSECKYFASA 755

RESULT 3

AAU72736

ID AAU72736 standard; Protein; 1729 AA.

XX AC AAU72736;

XX DT 26-FEB-2002 (first entry)

XX DE Arabidopsis Demeter protein.

XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;

XX KW DNA demethylation; transgenic plant; transcription modulation;

XX KW flowering time; endosperm development; MEDEA.

XX OS Arabidopsis Thaliana.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:08 ; Search time 35.7026 Seconds
(without alignments)
858.415 Million cell updates/sec

Title: US-09-840-743-72

Perfect score: 693

Sequence: 1 WXXXXXXXEXXXXXXX.....XXXXXXXXXXXXXXXLXXX 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	440	63.5	1413	23	Arabidopsis DMT2 (
2	433	62.5	1114	23	Arabidopsis DMT4 (
3	433	62.5	1729	23	Arabidopsis Demete
4	433	62.5	1952	23	Rice DMT1 protein.
5	432	62.3	298	23	Soybean DMT1 prote
6	432	62.3	1332	23	Arabidopsis DMT3 (
7	373.5	53.9	251	23	Soybean DMT2 prote
8	335.5	48.4	1309	23	Herbicidally activ
9	255	36.8	179	23	Soybean DMT protein
10	141	20.3	145	23	Wheat DMT2 protein

11	106	15.3	276	23	AAU72747	Corn DMT3 protein.
12	98	14.1	166	23	AAU72746	Corn DMT2 protein.
13	86	12.4	223	22	AAB96207	Putative P. abyss
14	74	10.7	218	23	ABB54387	Lactococcus lactis
15	72	10.4	219	23	ABB48015	Listeria monocytog
16	68	9.8	210	22	AAG81939	S. epidermidis ope
17	68	9.8	224	22	AAG82795	S. epidermidis ope
18	67	9.7	224	23	ABP38380	Staphylococcus epi
19	63	9.1	308	22	ABB58227	Drosophila melanog
20	61	8.8	1072	22	ABB69797	Drosophila melanog
21	59	8.5	350	19	AAW60244	Amino acid of the
22	59	8.5	350	22	AAU69755	Escherichia coli (
23	57.5	8.3	118	21	AAB37443	Human cystatin D.
24	57.5	8.3	122	21	AAAY81137	Human wild-type cy
25	57.5	8.3	122	21	AAAY81171	Human mutant cysta
26	57.5	8.3	122	21	AAAY81172	Human mutant cysta
27	57.5	8.3	122	21	AAAY81173	Human mutant cysta
28	57.5	8.3	122	21	AAAY81174	Human mutant cysta
29	57.5	8.3	122	21	AAAY81175	Human mutant cysta
30	57.5	8.3	122	21	AAAY81176	Human mutant cysta
31	57.5	8.3	122	21	AAAY81177	Human mutant cysta
32	57.5	8.3	122	21	AAAY81181	Human mutant cysta
33	57.5	8.3	122	21	AAAY81182	Human mutant cysta
34	57.5	8.3	142	22	AAE11210	Human cystatin D (
35	57.5	8.3	142	22	AAE04437	Human cystatin D h
36	57.5	8.3	142	22	AAE02408	Human cystatin D p
37	57.5	8.3	142	23	AAO15149	Human cystatin D p
38	56.5	8.2	122	21	AAAY81178	Human mutant cysta
39	56	8.1	195	21	AAG12675	Arabidopsis thalia
40	56	8.1	326	21	AAG12674	Arabidopsis thalia
41	55.5	8.0	320	19	AAW70220	Leishmania antigen
42	55.5	8.0	320	19	AAW70236	Leishmania antigen
43	55.5	8.0	320	23	ABG60872	Leishmania antigen
44	55.5	8.0	320	23	ABG60888	Leishmania antigen
45	55.5	8.0	320	23	AAU71816	Leishmania antigen

ALIGNMENTS

RESULT 1
AAU72737
ID AAU72737 standard; Protein; 1413 AA.
XX
AC AAU72737;
XX
DT 26-FEB-2002 (first entry)
XX
DE Arabidopsis DMT2 (IDMT2) protein.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Arabidopsis Thaliana.
XX
PN WO200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
DR WPI; 2002-055307/07.
XX
DR N-PSDB; AAS96692.
XX
PT New polynucleotide that control plant development comprising a sequence
PT having a specific homology to DEMETER domains A,B or C

Db 34 GYLTSSVRCVAGKSCPDVAVCKVL-GGSEFINNN----- 74
Qy 155 OKXLWPRCLKQXTLYELHYXXITFGKXFXCTKXXPNCNACPMXXXP 203
Db 75 -----CPCTDPNCASCPSDAGTC 92

RESULT 14

US-08-550-36
; Sequence 36, Application US/08808550
; Patent No. 5871992
; GENERAL INFORMATION:
; APPLICANT: Teebor, George W.
; APPLICANT: Hilbert, Timothy P.
; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,550
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-001 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
US-08-550-36

Query Match 7.0%; Score 48.5; DB 2; Length 207;
Best Local Similarity 23.4%; Pred. No. 5.9;
Matches 22; Conservative 7; Mismatches 38; Indels 27; Gaps 5;

Qy 103 ECVRLLLXXAFVDTNVGRIXVRXGVPLXPLPXXQXHLXLYPXXXXXQKXLWPR 162
Db 131 ECV-----GIAVDTHVHRISNRLGWIK-TSTPEKTQKALEILLP-----KSEW--- 172
Qy 163 CKLXQXTLYELHYXXITFGKXXFXCTKXXPNCNAC 196
Db 173 -----QPINHLLVGFQMQ-CQVPRPKGTC 197

RESULT 15

US-08-808-550-39
; Sequence 39, Application US/08808550
; Patent No. 5871992
; GENERAL INFORMATION:

; APPLICANT: Teebor, George W.
; APPLICANT: Hilbert, Timothy P.
; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,550
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-001 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 259 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: C. elegans
US-08-808-550-39

Query Match 7.0%; Score 48.5; DB 2; Length 259;
Best Local Similarity 23.4%; Pred. No. 7.4;
Matches 22; Conservative 7; Mismatches 38; Indels 27; Gaps 5;

Qy 103 ECVRLLLXXAFVDTNVGRIXVRXGVPLXPLPXXQXHLXLYPXXXXXQKXLWPR 162
Db 131 ECV-----GIAVDTHVHRISNRLGWIK-TSTPEKTQKALEILLP-----KSEW--- 172
Qy 163 CKLXQXTLYELHYXXITFGKXXFXCTKXXPNCNAC 196
Db 173 -----QPINHLLVGFQMQ-CQVPRPKGTC 197

Search completed: November 8, 2002, 19:29:16
Job time : 15.1536 secs

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183.861
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/022.765
FILING DATE: 12-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-183-861-55

Query Match 8.0%; Score 55.5; DB 4; Length 320;
Best Local Similarity 19.3%; Pred. No. 0.7;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDNNGRIVXRGXVPLXPLPXXXXXHXHLLXXYPXXXXX 154
Db 34 GYGLTSSVVCVRCVAGCKSCPVDANVCKVCL-GGSEPINNM----- 74
Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGKXXCTKXXPNACPMXXEC 203
Db 75 -----CPCDTPNCASCPSDAGTC 92

RESULT 12
US-09-022-765-22
Sequence 22, Application US/09022765
Patent No. 6375955
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.765
FILING DATE: 12-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-022-765-22
Query Match 8.0%; Score 55.5; DB 4; Length 320;
Best Local Similarity 19.3%; Pred. No. 0.7;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDNNGRIVXRGXVPLXPLPXXXXXHXHLLXXYPXXXXX 154
Db 34 GYGLTSSVVCVRCVAGCKSCPVDANVCKVCL-GGSEPINNM----- 74
Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGKXXCTKXXPNACPMXXEC 203
Db 75 -----CPCDTPNCASCPSDAGTC 92

RESULT 13
US-09-022-765-55
Sequence 55, Application US/09022765
Patent No. 6375955
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.765
FILING DATE: 12-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-022-765-55

Query Match 8.0%; Score 55.5; DB 4; Length 320;
Best Local Similarity 19.3%; Pred. No. 0.7;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDNNGRIVXRGXVPLXPLPXXXXXHXHLLXXYPXXXXX 154

Db 84 YYFNKFGRTT-CTKSQPNLNC 106

RESULT 9

US-09-241-376-4
; Sequence 4, Application US/09241376
; Patent No. 6300477
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ni, Jian
; APPLICANT: Rosen, Craig A.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Human Cystatin E
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/241,376
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/744,138
; FILING DATE: 05-NOV-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/461,030
; FILING DATE: 05-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF202PID1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 309 8504
; TELEFAX: 301 309 8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Cystatin D
US-09-241-376-4
Query Match 8.3%; Score 57.5; DB 4; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;
Qy 174 HYXXITFGKXXFCTKXXPNAC 197
;| :||: ||| || :||
Db 84 YYFNKFGRTT-CTKSQPNLNC 106

RESULT 10

US-09-183-861-22
; Sequence 22, Application US/09183861
; Patent No. 6365165
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/183,861
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/022,765
; FILING DATE: 12-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-183-861-22
Query Match 8.0%; Score 55.5; DB 4; Length 320;
Best Local Similarity 19.3%; Pred. No. 0.7;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDITNVGRXVRXGVPXLPLPXXXQXHLXLYPXXXXX 154
;| :||: ||| || :||
Db 34 GYGLTSSVVCVRCVAGCKSCPVDANVCVKCL-GGSEFINNM----- 74
Qy 155 QKXLWPLRLKXQXTLYELHYXXITFGKXXFXCTKXXPNACPMXXEC 203
;| :||: ||| || :||
Db 75 -----CPCTDPNCASCPDAGTC 92
RESULT 11
US-09-183-861-55
; Sequence 55, Application US/09183861
; Patent No. 6365165
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
; FILE REFERENCE: IL-10284
; CURRENT APPLICATION NUMBER: US/09/650,855
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 37
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-650-855-37

Query Match      8.5%; Score 59; DB 4; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.21;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps 3;

QY 96 GXGLKSECVRLXLXXXAPV-DTNGRIVRXGVPLPLPXXXXXHXLLXXYPXXXX 154
Db 116 GVGRSTAGAILSLGRH-FPILGDKVRVLAR-----CYAVSGWPGKKEV 160

QY 155 QKXLPRLCKLQXTLVELHYXITFGKXXFXCTKXXPNCNACPM 198
Db 161 ENKLWSLEQTPAVGVVERFNOAMMDLGAMICTRSKPKCSLCPL 204

RESULT 6
US-08-744-138-4
; Sequence 4, Application US/08744138
; Patent No. 6011012
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ni, Jian
; APPLICANT: Rosen, Craig A.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Human Cystatin E
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,138
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF202P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 309 8504
; TELEFAX: 301 309 8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Cystatin D
US-08-744-138-4
```

```
Query Match      8.3%; Score 57.5; DB 3; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 174 HYXXITFGKXXFCTKXXPCNACP 197
Db 84 YFNVKFGRTT-CTKSQPNLDNCP 106

RESULT 7
US-09-431-480-7
; Sequence 7, Application US/09431480
; Patent No. 6235708
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Feldhaus, Andrew
; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
; FILE REFERENCE: 98-72
; CURRENT APPLICATION NUMBER: US/09/431,480
; PRIOR FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: 60/109,217
; EARLIER FILING DATE: 1998-11-20
; EARLIER APPLICATION NUMBER: 60/156,382
; EARLIER FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-431-480-7

Query Match      8.3%; Score 57.5; DB 4; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 174 HYXXITFGKXXFCTKXXPCNACP 197
Db 84 YFNVKFGRTT-CTKSQPNLDNCP 106

RESULT 8
US-09-617-302-7
; Sequence 7, Application US/09617302
; Patent No. 6245529
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Feldhaus, Andrew
; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
; FILE REFERENCE: 98-72 C1
; CURRENT APPLICATION NUMBER: US/09/617,302
; CURRENT FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/431,480
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: 60/109,217
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: 60/156,382
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-617-302-7

Query Match      8.3%; Score 57.5; DB 4; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 174 HYXXITFGKXXFCTKXXPCNACP 197
Db 84 YFNVKFGRTT-CTKSQPNLDNCP 197
```

```
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,550
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-001 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus subtilis
; US-08-808-550-31

Query Match          9.2%; Score 64; DB 2; Length 209;
Best Local Similarity 23.1%; Pred. No. 0.02;
Matches 25; Conservative 10; Mismatches 55; Indels 18; Gaps 3;

QY 91 LXXGKLSXECVRLKXQXTLYELHYXXITFGKXXFCTKXXPNCNACPM 198
Db 112 LVKLGKGRKTNVVSVAFGPAIAVDTHVERVSKRLGICRWKDSVLEVEKTLMRKVP- 170
QY 151 XXXXOKLWPRCLKXQXTLYELHYXXITFGKXXFCTKXXPNCNACPM 198
Db 171 -----KEDWS-----VTHRLIFFGR-YHCKAQSPRCAECPL 201

RESULT 3
US-08-813-574-9
; Sequence 9, Application US/08813574
; Patent No. 6013473
; GENERAL INFORMATION:
; APPLICANT: Wei, Ying-Fei
; TITLE OF INVENTION: Human Muty
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,574
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/013,132
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-813-574-9

Query Match          8.5%; Score 59; DB 3; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.21;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps 3;

QY 96 GXGLKSXECVRLKXQXTLYELHYXXITFGKXXFCTKXXPNCNACPM 198
Db 116 GVGSRTAGAILSLSLGKH-FPILDGNVKKVLR-----CYAVSGWPGKKEV 160
QY 155 QKXLWPRCLKXQXTLYELHYXXITFGKXXFCTKXXPNCNACPM 198
Db 161 ENKLWSLSQVTPAVGVVERFNQAMMDLGAMICTRSPKCSLCPL 204

RESULT 4
US-09-651-656-37
; Sequence 37, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; POLYMORPHISMS, DNA DAMAGE AND DNA MISMATCHES
; TITLE OF INVENTION: POLYMORPHISMS, DNA DAMAGE AND DNA MISMATCHES
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-651-656-37

Query Match          8.5%; Score 59; DB 4; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.21;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps 3;

QY 96 GXGLKSXECVRLKXQXTLYELHYXXITFGKXXFCTKXXPNCNACPM 198
Db 116 GVGSRTAGAILSLSLGKH-FPILDGNVKKVLR-----CYAVSGWPGKKEV 160
QY 155 QKXLWPRCLKXQXTLYELHYXXITFGKXXFCTKXXPNCNACPM 198
Db 161 ENKLWSLSQVTPAVGVVERFNQAMMDLGAMICTRSPKCSLCPL 204

RESULT 5
US-09-650-855-37
; Sequence 37, Application US/09650855
; Patent No. 6365355
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	67	9.7	224	4	US-09-134-001C-3225	Sequence 3225, Appl
2	64	9.2	209	2	US-08-808-550-31	Sequence 31, Appl
3	59	8.5	350	3	US-08-813-574-9	Sequence 9, Appl
4	59	8.5	350	4	US-09-651-656-37	Sequence 37, Appl
5	59	8.5	350	4	US-09-650-855-37	Sequence 37, Appl
6	57.5	8.3	142	3	US-08-744-138-4	Sequence 4, Appl
7	57.5	8.3	142	4	US-09-431-480-7	Sequence 7, Appl
8	57.5	8.3	142	4	US-09-617-302-7	Sequence 7, Appl
9	57.5	8.3	142	4	US-09-241-376-4	Sequence 4, Appl
10	55.5	8.0	320	4	US-09-183-861-22	Sequence 22, Appl
11	55.5	8.0	320	4	US-09-183-861-55	Sequence 55, Appl
12	55.5	8.0	320	4	US-09-022-765-22	Sequence 22, Appl
13	55.5	8.0	320	4	US-09-022-765-55	Sequence 55, Appl
14	48.5	7.0	207	2	US-08-808-550-36	Sequence 36, Appl
15	48.5	7.0	259	2	US-08-808-550-39	Sequence 39, Appl
16	47.5	6.9	120	6	5432264-4	Patent No. 5432264
17	47.5	6.9	145	2	US-08-832-535-11	Sequence 11, Appl
18	47.5	6.9	146	2	US-08-791-522-3	Sequence 3, Appl
19	47.5	6.9	146	3	US-08-744-138-3	Sequence 3, Appl
20	47.5	6.9	146	3	US-09-019-485-4	Sequence 4, Appl
21	47.5	6.9	146	3	US-09-314-777-3	Sequence 3, Appl
22	47.5	6.9	146	4	US-09-431-480-6	Sequence 6, Appl
23	47.5	6.9	146	4	US-09-617-302-6	Sequence 6, Appl
24	47.5	6.9	146	4	US-09-241-376-3	Sequence 3, Appl
25	47.5	6.9	146	5	PCR-US95-07135-9	Sequence 9, Appl
26	47.5	6.9	146	6	5432264-6	Patent No. 5432264
27	47	6.8	191	3	US-08-974-022-52	Sequence 52, Appl

THIS PAGE BLANK (USPTO)

```
; PRIOR FILLING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 468
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-468

Query Match          6.6%; Score 45.5; DB 10; Length 494;
Best Local Similarity 32.4%; Pred No. 14;
Matches 11; Conservative 6; Mismatches 16; Indels 1; Gaps 1;

QY   104  CVRLXLXXAAPPDINVGRIXVRXGXVPLXP LP 137
      |::| :||| :| :|
Db    319  CAQMLAWSLPFYLQTVLGRSEVETGLL-LTPWP 351

RESULT 14
US-09-864-761-35736
; Sequence 35736, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PTOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35736
; LENGTH: 84
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011301.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EST_HUMAN HIT: AA306508.1, EVALUE 2.40e+00
US-09-864-761-35736

Query Match 6.5%; Score 45; DB 10; Length 84;
Best Local Similarity 50.0%; Pred. No. 2.4;
Matches 6; Conservative 3; Mismatches 3; Indels 0

QY 187 TKXXPNCNACPM 198
   :: ||| |||:
DB 33 SQAPNCTACPL 44

RESULT 15
US-09-775-932-6
; Sequence 6, Application US/09775932
; Patent No. US20020137671A1
; GENERAL INFORMATION:
; APPLICANT: University of British Columbia
; TITLE OF INVENTION: Production and use of Modified Cystatins
; FILE REFERENCE: 58069
; CURRENT APPLICATION NUMBER: US/09/775,932
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: CA99/00717
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: 60/095,503
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO.6
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-932-6

Query Match 6.5%; Score 45; DB 10; Length 121;
Best Local Similarity 27.8%; Pred. No. 3.6;
Matches 10; Conservative 6; Mismatches 20; Indels 0

QY 161 RLCKLQXLTLYELHXXKITFGKXXCTKXXPCNAC 196
   |::|:::||| ||| |||: |
DB 49 RVLRAQQTGVGVNYFDVEVGRTICTKSQPNLDT 84

Search completed: November 8, 2002, 19:29:41
Job time : 8.14052 secs

```

Search completed: November 8, 2002, 19:29:41
Job time : 8.14052 secs


```
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Calcium Channel Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT013PIC1
; CURRENT APPLICATION NUMBER: US/10/050,786
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 09/774,028
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: PCT/US00/20392
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/145,958
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 60/149,446
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: US 60/189,064
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 7
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-050-786-7
```

```
Query Match 7.3%; Score 50.5; DB 9; Length 584;
Best Local Similarity 36.1%; Pred. No. 2.6;
Matches 13; Conservative 2; Mismatches 20; Indels 1; Gaps 1;

QY 158 LMPRLCKLXQTYLHYHXXITFGKXXFCCKXXPNC 193
      |||: | | | | | | | | | | | | | | | |
Db 331 LLPRMASLG-LTLILFYYSFAIVGMEFFCGIVFPNC 365
```

```
RESULT 7
US-09-764-877-1188
; Sequence 1188, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1188
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1188
```

```
Query Match 6.9%; Score 47.5; DB 10; Length 82;
Best Local Similarity 22.2%; Pred. No. 0.91;
Matches 14; Conservative 10; Mismatches 34; Indels 5; Gaps 2;
```

```
QY 116 PVDTNVGRIXVRXGVPLXPLPXXXXXHXLPXXXXXQKXLPRLCKLXQTYLHY 175
      | | | | | | | | | | | | | | | | | |
Db 20 PSPHTASRLPIR----PFLVITAYGD-HILRFFPAQDSEDFSLYPMSCVIHYITLFLV 74

QY 176 XXI 178
;
Db 75 FSV 77
```

```
RESULT 8
US-09-775-932-2
; Sequence 2, Application US/09775932
; Patent No. US20020137671A1
; GENERAL INFORMATION:
; APPLICANT: University of British Columbia
; TITLE OF INVENTION: Production and use of Modified Cystatins
; FILE REFERENCE: 58069
```

```
; CURRENT APPLICATION NUMBER: US/09/775,932
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: CA99/00717
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: 60/095,503
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-932-2
```

```
Query Match 6.9%; Score 47.5; DB 10; Length 120;
Best Local Similarity 33.3%; Pred. No. 1.4;
Matches 8; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
```

```
QY 174 HYXXITFGKXXFCCKXXPNCNACP 197
      :: | | | | | | | | | | | |
Db 62 YFLDVELGRTT-CTKTQPNLDNCP 84
```

```
RESULT 9
US-09-940-497-3
; Sequence 3, Application US/09940497
; Patent No. US20020052476A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Human Cystatin E
; FILE REFERENCE: PF202FID2
; CURRENT APPLICATION NUMBER: US/09/940,497
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 09/241,376
; PRIOR FILING DATE: 1999-02-02
; PRIOR APPLICATION NUMBER: US 08/744,138
; PRIOR FILING DATE: 1996-11-05
; PRIOR APPLICATION NUMBER: US 08/461,030
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-940-497-3
```

```
Query Match 6.9%; Score 47.5; DB 10; Length 146;
Best Local Similarity 33.3%; Pred. No. 1.7;
Matches 8; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
```

```
QY 174 HYXXITFGKXXFCCKXXPNCNACP 197
      :: | | | | | | | | | | | |
Db 88 YFLDVELGRTT-CTKTQPNLDNCP 110
```

```
RESULT 10
US-09-969-834-3
; Sequence 3, Application US/09969834
; Patent No. US20020102711A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga K.
; TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
```

; PRIOR FILING DATE: 1999-02-02
; PRIOR APPLICATION NUMBER: US 08/744,138
; PRIOR FILING DATE: 1996-11-05
; PRIOR APPLICATION NUMBER: US 08/461,030
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-940-497-4

Query Match 8.3%; Score 57.5; DB 10; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.037;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Qy 174 HYXXITFGKXXFCFKXXPNACNP 197
Db 84 YFNVKFGRTT-CTKSPNLDNCP 106

RESULT 3

US-09-874-923-22
; Sequence 22, Application US/09874923
; Patent No. US20020081320A1

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Leishmania major
US-09-874-923-22

Query Match 8.0%; Score 55.5; DB 10; Length 320;
Best Local Similarity 19.3%; Pred. No. 0.2;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;

Qy 96 GXGLKSXE-CVRLLLXXXAFVDTNVGRXGXVPLPPLPXXXXHXLXXYPXXXXX 154
Db 34 GYGLTSSVRCVRCVAGCKPCPDANVCKVCL-GGSEPINNM----- 74

Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGKXXFCFKXXPNACNPXXXXXEC 203
Db 75 -----CPCTDPNCACSPSDAGTC 92

RESULT 4

US-09-874-923-55
; Sequence 55, Application US/09874923
; Patent No. US20020081320A1

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Leishmania major
US-09-874-923-55

Query Match 8.0%; Score 55.5; DB 10; Length 320;
Best Local Similarity 19.3%; Pred. No. 0.2;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;

Qy 96 GXGLKSXE-CVRLLLXXXAFVDTNVGRXGXVPLPPLPXXXXHXLXXYPXXXXX 154
Db 34 GYGLTSSVRCVRCVAGCKPCPDANVCKVCL-GGSEPINNM----- 74

Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGKXXFCFKXXPNACNPXXXXXEC 203
Db 75 -----CPCTDPNCACSPSDAGTC 92

RESULT 5

US-09-874-923-121
; Sequence 121, Application US/09874923
; Patent No. US20020081320A1

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Leishmania major and chagasi
US-09-874-923-121

Query Match 8.0%; Score 55.5; DB 10; Length 709;
Best Local Similarity 19.3%; Pred. No. 0.47;
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;

Qy 96 GXGLKSXE-CVRLLLXXXAFVDTNVGRXGXVPLPPLPXXXXHXLXXYPXXXXX 154
Db 423 GYGLTSSVRCVRCVAGCKPCPDANVCKVCL-GGSEPINNM----- 463

Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGKXXFCFKXXPNACNPXXXXXEC 203
Db 464 -----CPCTDPNCACSPSDAGTC 481

RESULT 6

US-10-050-786-7
; Sequence 7, Application US/10050786
; Patent No. US20020155539A1

THIS PAGE BLANK (USPTO)

```

DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III.1; 1.
DR PROSITE; PS01155; ENDONUCLEASE_III.2; 1.
KW Hypothetical protein; DNA repair; Hydrolase; Glycosidase; Lyase;
KW Iron-sulfur; 4Fe-4S.
FT METAL 187 187 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 194 194 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 197 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 203 203 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 259 AA; 28841 MW; F52E57810571EDBC CRC64;

Query Match 7.0%; Score 48.5; DB 1; Length 259;
Best Local Similarity 23.4%; Pred No. 2,9;
Matches 22; Conservative 7; Mismatches 38; Indels 27; Gaps

QY 103 ECVRLRLXXXXAPPVDNTVGRIXVRXGXVPLXPLPXXQXHXLXXYPXXXXXQKXLWPL 162
||| ||| ||| ||| ||| ||| |||
Db 131 ECV-----GIAVDTHVHRISNLGMWK-TSTPEKTQKALEILLP-----KSEW--- 172

QY 163 CKLXQTYLHYELHYXXITFGKXXFCTKXXPNMCAC 196
::: ||| ||| |||
Db 173 -----QPINHLLVGFQMGQ-CQPVRPKCGTC 197

```

Search completed: November 8, 2002, 19:27:11
Job time : 10.268 secs

```
RA Lin C.-S.;
RT "Nucleotide sequence of the essential region of bacteriophage P4.";
RL Nucleic Acids Res. 12:8667-8684(1984).
CC -!- SIMILARITY: SOME, TO E. COLI PROPHAGE CP4-57 REGULATORY PROTEIN
CC ALPA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51522; CAA35903.1; -.
CC DR EMBL; X02534; CAA26377.1; -.
CC DR PIR; G23878; QJBP4.
CC DR PIR; J00029; J00029.
CC KW Hypothetical protein; DNA-binding.
CC FT DNA_BIND 30 49 H-T-H MOTIF (POTENTIAL).
CC SQ SEQUENCE 88 AA; 9749 MW; DA16EF6E8B7BD70F CRC64;
CC -----
Query Match 7.1%; Score 49.5; DB 1; Length 88;
Best Local Similarity 33.3%; Pred. No. 0.69; Indels 15; Gaps 2;
Matches 15; Conservative 2; Mismatches 13;
QY 132 PLXPLPXXQXHLXLYPXXXXXQKXLP-----RLCKLXQXTLYEL 173
||| ||| | | | | | | | | | | | | | | | |
Db 15 PLMLPDITQERLVR-----PEVMHLCGLSRSTIYEL 47
-----
RESULT 14
YDK9 SCHPO STANDARD; PRT; 1033 AA.
AC P87L15;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C20G8.09c in chromosome I.
GN SPAC20G8.09c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds R., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vansteelt E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipkovski G.V., Ussery D., Barrell B.G., Nurse P.;
```

```
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE UPF0202 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z95334; CAB08603.1; -.
CC DR EMBL; X02534; CAA26377.1; -.
CC FT NP_BIND 282 289 ATP (POTENTIAL).
CC SQ SEQUENCE 1033 AA; 116463 MW; 8432B313D818E135 CRC64;
CC -----
Query Match 7.1%; Score 49; DB 1; Length 1033;
Best Local Similarity 27.9%; Pred. No. 9.1;
Matches 12; Conservative 6; Mismatches 25; Indels 0; Gaps 0;
QY 156 KXLPRLCKLXQXTLYELHYXXITFGKXXCTKXXPCNACPM 198
| : | | | | : | | | | | | | | | | | | | | | |
Db 691 KTMPPLLLKLELOPEPLHYGVSYGLTPSLQKFKWREGYCP 733
-----
RESULT 15
END3 CAEEL STANDARD; PRT; 259 AA.
AC P34137;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable endonuclease III homolog (EC 4.2.99.18) (DNA-(Apurinic or
DE apyrimidinic site) lyase).
GN R10E4.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RL Ainscough R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE
CC ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -!- COPACATOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z50874; CAA90766.1; -.
CC DR EMBL; X02534; CAA26377.1; -.
CC DR HSSP; P20625; 2ABK.
CC DR WormPep; R10E4.5; CE03559.
CC DR InterPro; IPR004035; EndoIII_FCL.
CC DR InterPro; IPR004036; EndoIII_Hhh.
CC DR InterPro; IPR003265; Endo_3c.
CC DR InterPro; IPR003651; Res_bind.
CC DR Pfam; PF00730; Hhh-GPD; 1.
```

```
Matches 16; Conservative 6; Mismatches 30; Indels 1; Gaps 1;

Qy 76 LEWLRRXXXXXXKXLLXXGXGLKSECVRLXLX-XAFFPVDVTNGRIXVR 127
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 225 LQQLRESSYEAAHRCALCILPGVGTQVADICIMALDKRQAPVVDVHMWHIAQR 277

RESULT 11
ID OGG1_METTH STANDARD; PRT; 312 AA.
AC O27397;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable N-glycosylase/DNA-lyase [Includes: 8-oxoguanine DNA
DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
DE (EC 4.2.99.18) (AP lyase)].
GN MTH1342.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: DNA REPAIR ENZYME THAT INCISES DNA AT 8-OXOG RESIDUES.
CC EXCISES 7,8-DIHYDRO-8-OXOGUANINE AND 2,6-DIAMINO-4-HYDROXY-5-N-
CC METHYLFORMAMIDOPYRIMIDINE (FAPY) FROM DAMAGED DNA. HAS A BETA-
CC LYASE ACTIVITY THAT NICKS DNA 3' TO THE LESION (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage near apurinic or
CC apyrimidinic sites to products with 5'-phosphate.
CC -!- SIMILARITY: BELONGS TO THE OGG1 FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000897; AAB85820.1; -.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR004577; Ogg.
DR Pfam; PF00730; HhH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00278; HhH1; 1.
DR TIGRFAMs; TIGR00588; Ogg; 1.
KW Hydrolase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
KW Multifunctional enzyme; Complete proteome.
FT ACT_SITE 235 235 BY SIMILARITY.
SQ SEQUENCE 312 AA; 36051 MW; D33B4FF497BC5A1C CRC64;

Query Match 7.4%; Score 51.5; DB 1; Length 312;
Best Local Similarity 40.08; Pred. No. 1.1;
Matches 14; Conservative 3; Mismatches 17; Indels 1; Gaps 1;

Qy 91 LLXXGXGLKSECVRLXLX-XXAFPPVDVTNGRI 124
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 226 LLELGGVGPVKVADICILLYGFRKTEAFPVDVWIRI 260

RESULT 12
YNN2_YEAST STANDARD; PRT; 1056 AA.
ID P53914;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 119.3 kDa protein in FPR1-TOM22 intergenic region.
GN YNL132W OR N1216 OR N1858.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96109932; PubMed=8619318;
RA Mallet L., Bussereau F., Jacquet M.;
RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT MEP2, CAP/SRV2, NAM9, FKBI/FPR1/RBP1, MOM22 and CPT1, predicts an
RT adenosine deaminase gene and 14 new open reading frames.";
RL Yeast 11:1195-1209(1995).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE UPF0202 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46843; CAA86893.1; -.
DR EMBL; Z71408; CAA96014.1; -.
DR SGD; S0005076; YNL132W.
KW Hypothetical protein; ATP-binding; Nuclear protein.
FT NP_BIND 283 290 ATP (POTENTIAL).
SQ SEQUENCE 1056 AA; 119347 MW; 76721ED0867ED618 CRC64;

Query Match 7.4%; Score 51; DB 1; Length 1056;
Best Local Similarity 30.2%; Pred. No. 4.3;
Matches 13; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

Qy 156 KXLWPRLCXQXLYELHYYXITFGKXXCTCXXXPCNACPM 198
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 703 KTLPLLLKLSEOPPHLYLGVSYGLTQSLHKEFWKNSFVPV 745

RESULT 13
Y9K_BPP4 STANDARD; PRT; 88 AA.
ID P12552;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Hypothetical 9.7 kDa protein (ORF88) (Putative DNA-binding protein).
OS Bacteriophage P4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC unclassified Myoviridae.
OX NCBI_TaxID=10680;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90221913; PubMed=2183201;
RA Halling C., Calendar R., Christle G.E., Dale E.C., Deho G., Finkel S.,
RA Flensburg J., Ghisotti D., Kahn M.L., Lane K.B., Lin C.-S.,
RA Lindqvist B.H., Pierson L.S., Six E.W., Sunshine M.G., Ziermann R.;
RT "DNA sequence of satellite bacteriophage P4.";
RL Nucleic Acids Res. 18:1649-1649(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85062840; PubMed=6095206;
```


RESULT 9
MUTY_SALTY STANDARD; PRT; 350 AA.
ID MUTY_SALTY
AC Q05869;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE A/G-specific adenine glycosylase (EC 3.2.2.-).
GN MUTY OR MUTB OR STM3110.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GW1803;
RX MEDLINE=93123173; PubMed=9419300;
RA Desiraju V., Shanabru W.G., Lu A.L.;
RT "Nucleotide sequence of the Salmonella typhimurium mutB gene, the
homolog of Escherichia coli mutY.";
RL J. Bacteriol. 175:541-543(1993).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=2153498; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISPAIRS.
CC -!- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86634; AAA27165.1; -;
DR EMBL; AE008843; AAL21985.1; -;
DR PIR; A40647; A40647.
DR HSSP; P17802; 1MUN.
DR StyGene; SG10238; mutY.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR004036; EndoIII_Hh.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; Fes_bind.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00730; HhH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
DR SMART; SM00278; Hhh1; 1.
DR TIGRFAMs; TIGR01084; mutY; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
KW DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;
Complete proteome.
FT METAL 192 192 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 199 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 202 202 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 208 208 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 208 208 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 350 AA; 39409 MW; 5C55DB4D7B7BB69F CRC64;

Query Match 7.9%; Score 55; DB 1; Length 350;
Best Local Similarity 19.5%; Pred. No. 0.33;
Matches 22; Conservative 12; Mismatches 45; Indels 34; Gaps 4;
QY 96 GXLKSECVRLLLXLLXXAFVDTNVGRGXVPLXPLPXXHXHXXYPXXXXXO 155
DB 116 GVGRTAGAILSLALGKHYPILDGNVKKVRLAR-----CYAVSWGPKKEVE 161
QY 156 KXLWPRCLCKLXQXTLYE-----LHYXXITFCXXXCTKXXPNACPM 198
DB 162 NTLW-----TLSEQVTPARGVERENQAMMDLG-AMVCTRSKPKCTLCPL 204
RESULT 10
OGGI_HUMAN STANDARD; PRT; 345 AA.
AC O15527; Q00390; P78554; Q00670; Q00705; O14876; Q95488; Q9ULJ34;
AC Q9Y6C3; Q9Y6C4; Q9UIK0; Q9UIK1; Q9UIK2; Q9Y2C0; Q9Y2C1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE N-glycosylase/DNA lyase [Includes: 8-oxoguanine DNA glycosylase
(EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
(EC 4.2.99.18) (AP lyase)].
GN OGG1 OR MMH OR MUTM OR OGH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX TISSUE=Colon;
RX MEDLINE=97330655; PubMed=9187114;
RA Aburatani H., Hippo Y., Ishida T., Takashima R., Matsuba C.,
RA Kodama T., Takao M., Yasui A., Yamamoto K., Asano M., Fukasawa K.,
RA Yoshinari T., Inoue H., Otsuka E., Nishimura S.;
RT "Cloning and characterization of mammalian 8-hydroxyguanine-specific
DNA glycosylase/apurinic, apyrimidinic lyase, a functional mutM
homologue.";
RL Cancer Res. 57:2151-2156(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1A).
RX MEDLINE=97352815; PubMed=9207108;
RA Rosengquist T.A., Zharkov D.O., Grollman A.P.;
RT "Cloning and characterization of a mammalian 8-oxoguanine DNA
glycosylase.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:7429-7434(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=97368311; PubMed=9223306;
RA Roldan-Arjona T., Wei Y.-F., Carter K.C., Klungland A., Anselmino C.,
RA Wang R.-P., Augustus M., Lindahl T.;
RT "Molecular cloning and functional expression of a human cDNA encoding
the antitumor enzyme 8-hydroxyguanine-DNA glycosylase.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8016-8020(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1A).
RX MEDLINE=97368310; PubMed=9223305;
RA Radicella J.P., Dherin C., Desmaze C., Fox M.S., Boiteux S.;
RT "Cloning and characterization of hOGG1, a human homolog of the OGG1
gene of Saccharomyces cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8010-8015(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1A).
RX MEDLINE=97342862; PubMed=9197244;
RA Lu R., Nash H.M., Verdine G.L.;
RT "A mammalian DNA repair enzyme that excises oxidatively damaged
guanines maps to a locus frequently lost in lung cancer.";
RL Curr. Biol. 7:397-407(1997).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 1A).
RX MEDLINE=97334205; PubMed=9190902;

[illegible]

Query Match	8.2%	Score 56.5	DB 1	Length 213
Best Local Similarity	25.0%	Pred. No. 0.11		
Matches	28	Conservative	8	Mismatches 45; Indels 31; Gaps 4;
QY	91	ILXXGXGLKXECVRLILXXAFPDVNTGRIYVRGXVPLDPLPXXXQXHLXXYPX	150	
Db	104	LKLPGVGRKTANIVLWVGFKKPALAVDTHVHRSNRLGW	144	
QY	151	XXXXOKXLWPR-----LCKLXOXTLY-ELHYXXITFGKXXFCTKXXPNMCNAC	196	
Db	145	-----KRTPTETEALKKLIPEDLWGPINGSWVEFGR-RICKPONPLCEEC	190	
RESULT 6				
OGGI_RAT	ID	OGGI_RAT	STANDARD;	PRT; 345 AA.
AC	070249;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	N-glycosylase; DNA lyase [Includes: 8-oxoguanine DNA glycosylase			
DE	(EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase			
DE	(EC 4.2.99.18) (AP lyase)]			
DN	OGGI OR NMH OR OGGI.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Hepatoma;			
RC	MEDLINE=99030527; PubMed=9801319;			
RT	Prieto Alam M.J., Jurado J., Francastel E., Laval F.;			
RT	"Rat 7,8-dihydro-8-oxoguanine DNA glycosylase: substrate specificity,			
RT	kinetics and cleavage mechanism at an apurinic site.";			
RL	Nucleic Acids Res. 26:5199-5202(1998).			
CC	!- FUNCTION: DNA REPAIR ENZYME THAT INCISES DNA AT 8-OXOG RESIDUES.			
CC	EXCISES 7,8-DIHYDRO-8-OXOGUANINE AND 2,6-DIAMINO-4-HYDROXY-5-N-			
CC	METHYLFORMAMIDOPYRIMIDINE (FAPY) FROM DAMAGED DNA. HAS A BETA-			
CC	LYASE ACTIVITY THAT NICKS DNA 3' TO THE LESION.			
CC	!- CATALYTIC ACTIVITY: Endonucleolytic cleavage near apurinic or			
CC	apyrimidinic sites to products with 5'-phosphate.			
CC	!- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC	!- SIMILARITY: BELONGS TO THE OGGI FAMILY 1.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed, usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; AF029690; AAC77525.1; "			
DR	InterPro; IPR003265; Endo_3c.			
DR	InterPro; IPR004577; Ogg.			
DR	Pfam; PF007730; Hhh-GPD; 1.			
DR	SMART; SM00478; ENDO3c; 1.			
DR	TIGRfams; TIGR00588; ogg; 1.			
DR	Hydrolase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;			
KW	Multifunctional enzyme; Nuclear protein.			
FT	ACT_SITE 249 249 BY SIMILARITY.			
FT	SEQUENCE 345 AA; 38711 MW; B7FDF8C782644C41 CRC64;			
QY	92	ILXXGXGLKXECVRLILXXAFPDVNTGRIYVRGXVPLDPLPXXXQXHLXXYPX	150	
Db	104	LKLPGVGRKTANIVLWVGFKKPALAVDTHVHRSNRLGW	144	
QY	151	XXXXOKXLWPR-----LCKLXOXTLY-ELHYXXITFGKXXFCTKXXPNMCNAC	196	
Db	145	-----KRTPTETEALKKLIPEDLWGPINGSWVEFGR-RICKPONPLCEEC	190	

ENBL; U28377; AAA69128.1; -
ENBL; AE000378; AAC75998.1; -
DR PIR; JQ0546; JQ0546.
DR PIR; B38535; B38535.
DR PDB; 1MUD; 29-OCT-99.
DR PDB; 1MUN; 26-AUG-99.
DR PDB; 1MU; 20-AUG-99.
DR EcoGene; EG10627; mutY.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR004036; EndoIII_HH.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; FeS_bind.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00730; HhH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; PES; 1.
DR SMART; SM00278; HHH1; 1.
DR TIGRFS; TIGR01084; mutY; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
KW DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S; 3D-structure
KW Complete proteome.
FT METAL 192 192 IRON-SULFUR (4FE-4S).
FT METAL 199 199 IRON-SULFUR (4FE-4S).
FT METAL 202 202 IRON-SULFUR (4FE-4S).
FT METAL 208 208 IRON-SULFUR (4FE-4S).
SQ SEQUENCE 350 AA; 39149 MW; C7D3657C03EBBF4F CRC64;

Query Match 8.5%; Score 59; DB 1; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.071;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps

QY 96 GXLGKXECVRLRLXLLXXXAPPV-DTNVGRIXVRXGXVPLXPXXXXXHXLLXXVPPXXXX 154
DB 116 GVRSTAGAILSLSLGKH-FPLDGNVXKVLAR-----CYVSGWFGKKEV 160

QY 155 QKXLWRLPCLKXOXTLYELHYXITFGKXXXFTKXXPNACPM 198
DB 161 ENKLWSLSQVTPAVGVERFNQAMMDLGAMICTRSKPKSLCPL 204

RESULT 4
CYTD_HUMAN
ID CYTD_HUMAN STANDARD; PRT; 142 AA.
AC P28325;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cystatin D precursor.
GN CST5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041895; PubMed=1939105;
RA Freije J.P., Abrahamson M., Olafsson I., Velasco G., Grubb A.,
Lopez-Otin C.;
RT "Structure and expression of the gene encoding cystatin D, a novel
human cysteine proteinase inhibitor.";
RL J. Biol. Chem. 266:20538-20543(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Saliva;
RC MEDLINE=93340179; PubMed=8340398;
RA Freije J.P., Balbin M., Abrahamson M., Velasco G., Grubb A.,
Lopez-Otin C.;
RT "Human cystatin D. cDNA cloning, characterization of the Escherichia
coli expressed inhibitor, and identification of the native protein in
saliva.";
RL J. Biol. Chem. 268:15737-15744(1993).
RN [3]

RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -!- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U11289; AAA80005.1; -;
DR EMBL; L47709; AAB38457.1; -;
DR EMBL; Z99115; CAB14150.1; -;
DR HSSP; P20625; 2ABK.
DR Subtilisin; BG10956; nth.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR004036; EndoIII_Hhh.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; Fes_bind.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00730; Hhh-GPD; 1.
DR SMART; SM00478; Endo3c; 1.
DR SMART; SM00525; FES; 1.
DR SMART; SM00278; Hhh1; 1.
DR TIGRfams; TIGR01083; nth; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;
KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 189
FT METAL 196 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 205 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 219 AA; 25000 MW; C38C66DA7948BA40 CRC64;

Query Match 9.2%; Score 64; DB 1; Length 219;
Best Local Similarity 23.1%; Pred. No. 0.0068;
Matches 25; Conservative 10; Mismatches 55; Indels 18; Gaps 3;

Qy 91 LXXXXXGKSKXECVRLXLLXXXXAFPDVTNVGRITXVRXGVPLPLPXXXXXHLXXYPX 150
Db 112 LVKLPVGGRKTNVVSVAFGVPAIVDTHVERVSKRIGICRWKDSVLEVEKTLMRKVP- 170
Qy 151 XXXXQKXLPRLCKLXQXTLYELAHYXXITFGKXXFXCTKXXPCNACPM 198
Db 171 -----KEDWS-----VTHRLIFFGR-YHCKAQSPRCAECPL 201

RESULT 2
END3_RICPR
ID_END3_RICPR STANDARD; PRT; 212 AA.
AC 005956;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site)
DE lyase).
GN NTH OR RP746.

OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RC MEDLINE=97419517; PubMed=9274032;
RA Andersson J.O., Andersson S.G.E.;
RT "Genomic rearrangements during evolution of the obligate
RT intracellular parasite Rickettsia prowazekii as inferred from an
RT analysis of 52015 bp nucleotide sequence";
RL Microbiology 143:2783-2795(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RC MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sierseitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria";
RL Nature 396:133-140(1998).
CC -!- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE
CC ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. INCISES DAMAGED DNA AT
CC CYTOSINES, THYMINE AND GUANINES. ACTS ON A DAMAGED STRAND, 5'
CC FROM THE DAMAGED SITE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -!- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11778; CAA72458.1; -;
DR EMBL; AJ235273; CAA15174.1; -;
DR HSSP; P20625; 2ABK.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR004036; EndoIII_Hhh.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; Fes_bind.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00730; Hhh-GPD; 1.
DR SMART; SM00478; Endo3c; 1.
DR SMART; SM00525; FES; 1.
DR SMART; SM00278; Hhh1; 1.
DR TIGRfams; TIGR01083; nth; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;
KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 187
FT METAL 187 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 194 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 203 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 212 AA; 24174 MW; A7FDA4F6033A6382 CRC64;

Query Match 9.2%; Score 63.5; DB 1; Length 212;
Best Local Similarity 24.1%; Pred. No. 0.0079;
Matches 26; Conservative 10; Mismatches 53; Indels 19; Gaps 3;

Qy 91 LXXXXXGKSKXECVRLXLLXXXXAFPDVTNVGRITXVRXGVPLPLPXXXXXHLXXYPX 150
Db 112 LVKLPVGGRKTNVVSVAFGVPAIVDTHVERVSKRIGICRWKDSVLEVEKTLMRKVP- 170
Qy 151 XXXXQKXLPRLCKLXQXTLYELAHYXXITFGKXXFXCTKXXPCNACPM 198
Db 171 -----KEDWS-----VTHRLIFFGR-YHCKAQSPRCAECPL 201

RESULT 2
END3_RICPR
ID_END3_RICPR STANDARD; PRT; 212 AA.
AC 005956;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site)
DE lyase).
GN NTH OR RP746.

THIS PAGE BLANK (USPTO)

```
FT REPEAT 2374 2472 CADHERIN 22.
FT REPEAT 2481 2576 CADHERIN 23.
FT REPEAT 2586 2686 CADHERIN 24.
FT REPEAT 2695 2793 CADHERIN 25.
FT REPEAT 2804 2897 CADHERIN 26.
FT REPEAT 2927 3020 CADHERIN 27.
FT VARIANT 1131 1153 MISSING (IN ALLELE D1).
FT CONFLICT 1070 1070 I -> V (IN REF. 3).
FT CONFLICT 1490 1490 S -> R (IN REF. 3).
FT CONFLICT 1636 1636 S -> G (IN REF. 3).
FT CONFLICT 1692 1692 P -> S (IN REF. 3).
FT CONFLICT 1804 1804 I -> V (IN REF. 3).
FT CONFLICT 2029 2029 I -> L (IN REF. 3).
FT CONFLICT 2210 2210 A -> P (IN REF. 3).
FT CONFLICT 2289 2289 S -> A (IN REF. 3).
FT CONFLICT 2536 2536 T -> S (IN REF. 3).
FT CONFLICT 2862 2862 Q -> R (IN REF. 3).
FT CONFLICT 3038 3038 G -> S (IN REF. 3).
SQ SEQUENCE 3503 AA; 379721 MW; F256924D5FB5F7AE CRC64;
```

Query Match 6.3%; Score 37; DB 5; Length 3503;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

```
QY 120 DXHXLXXDXLXXLXIXWP 140
Db 388 DGHFALTTRDNSIYLVHLP 408
```

Search completed: November 8, 2002, 19:28:42
Job time : 40.2614 secs

```

RA RP SEQUENCE FROM N.A.
RA Lennard N.;
RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RA [2]
RA MEDLINE=99069613; PubMed=9851916;
RX SEQUENCE FROM N.A.
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z75552; CAA99942.1; -
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 1.
KW DNA-binding; Zinc-finger
SQ SEQUENCE 624 AA; 71971 MW; 39E9BD2838A9C8D3 CRC64;

Query Match 6.3%; Score 37; DB 5; Length 624;
Best Local Similarity 64.3%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 109 LRTEHXYXLPDXH 122
|| || | || |
Db 90 LRIEHDV--LPDVH 101

RESULT 15
Q24292 PRELIMINARY; PRT; 3503 AA.
ID Q24292
AC Q24292; Q9VPS4;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DR 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DACHSOUS protein precursor (ADHERIN).
GN DS OR CG17941.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=EMBRYO;
RX MEDLINE=95324813; PubMed=7601355;
RA Clark H.F., Brentrup D., Schneitz K., Bieber A., Goodman C., Noll M.;
RT "Dachsous encodes a member of the cadherin superfamily that controls
RT imaginal disc morphogenesis in Drosophila.";
RL Genes Dev. 9:1530-1542(1995).
RN [2]
RP REVISIONS.
RA Noll M.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burdits K.C., Busan M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: INVOLVED IN MORPHOGENESIS. MAY ALSO BE INVOLVED IN CELL
CC ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN EMBRYONIC ECTODERM. IN LARVAE,
CC EXPRESSION IS RESTRICTED TO IMAGINAL DISKS AND BRAIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS WHERE IT
CC IS FIRST DETECTED DURING GASTRULATION. ALSO EXPRESSED IN LARVAE
CC AND ADULTS.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR EMBL: L08811; AAA79129.2; -
DR EMBL: AE003588; AAF51468.2; -
DR HSPD: P15116; INCGJ.
DR FlyBase: FBgn0000497; ds.
DR InterPro: IPR002126; Cadherin.
DR InterPro: IPR00233; Cadherin_C-term.
DR Pfam: PF00028; cadherin; 26.
DR Pfam: PF01049; Cadherin_C-term; 1.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 26.
DR PROSITE: PS00232; CADHERIN_1; 19.
DR PROSITE: PS00268; CADHERIN_2; 27.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Polymorphism; Signal
FT SIGNAL 1
FT CHAIN 21 3503
FT DOMAIN 21 3045
FT TRANSMEM 3046 3066
FT DOMAIN 3067 3503
FT REPEAT 58 121
FT REPEAT 130 233
FT REPEAT 242 340
FT REPEAT 353 451
FT REPEAT 460 558
FT REPEAT 567 662
FT REPEAT 671 774
FT REPEAT 783 878
FT REPEAT 887 983
FT REPEAT 992 1100
FT REPEAT 1109 1203
FT REPEAT 1213 1311
FT REPEAT 1320 1431
FT REPEAT 1440 1548
FT REPEAT 1563 1665
FT REPEAT 1674 1793
FT REPEAT 1803 1899
FT REPEAT 1907 2003
FT REPEAT 2012 2146
FT REPEAT 2122 2269
FT REPEAT 2261 2357

```

```

OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
[1]
RP SEQUENCE FROM N.A.
RA Murphy L., Harris D.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler S.P., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL035480; CAB36627.1; -.
DR EMBL; AL583925; CAC31872.1; -.
DR Leproma; ML2356; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR Phosphopantetheine; Transferase; Complete proteome.
KW Phosphopantetheine; Transferase; Complete proteome.
SQ SEQUENCE 1540 AA; 164209 MW; D7042CAAE672375 CRC64;

Query Match 6.4%; Score 38; DB 16; Length 1540;
Best Local Similarity 36.4%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

Oy ~ 112 EHXYVXLPDXHXXLXXXXXX 133
:| | | | | | | | | | | | | |
Db 990 DHDVSTLHDLHNLRGVDNVLY 1011

RESULT 12
Q92SK4
ID Q92SK4 PRELIMINARY; PRT; 214 AA.
AC Q92SK4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein R00374.
GN R00374 OR SMC01146.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gouzie T., Goffeau A., Kahn D., Kiss E., Leilaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RT Sinorhizobium meliloti strain 1021;
OC Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591783; CAC41811.1; -.

```

```

DR InterPro; IPR002637; Hamlp_like.
DR Pfam; PF01725; Hamlp_like; 1.
DR TIGRFAMS; TIGR00042; Hamlp_like; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 23109 MW; A282C7539CBBCDD CRC64;

Query Match 6.3%; Score 37; DB 16; Length 214;
Best Local Similarity 31.4%; Pred. No. 7.1;
Matches 11; Conservative 3; Mismatches 15; Indels 6; Gaps 2;

Oy 110 RTEHXVXXL---PDHXXLXXXXXXLLXIWP 140
| | | | | | | | | | | | | |
Db 125 RTAREVSVLCCLAWPDGHVELRGEVGV--VWPP 157

RESULT 13
Q966M0
ID Q966M0 PRELIMINARY; PRT; 325 AA.
AC Q966M0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein C55B7.11.
GN C55B7.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wamsley P.;
RT "The sequence of C. elegans cosmid C55B7.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006625; AAK68271.1; -.
SQ SEQUENCE 325 AA; 37141 MW; 55EAA4186BC645C0 CRC64;

Query Match 6.3%; Score 37; DB 5; Length 325;
Best Local Similarity 36.0%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

Oy 109 LRTEHXVXXLPDXHXXLXXXXXX 133
| | | | | | | | | | | | | |
Db 140 LNTETLIEILPDGKILNPRDASKY 164

RESULT 14
Q23159
ID Q23159 PRELIMINARY; PRT; 624 AA.
AC Q23159;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE W04D2.4 protein.
GN W04D2.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;

```

```

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF408195; AAL0551.1; -.
DR InterPro: IPR002934; NTP_transf.
DR Pfam: PF01909; NTP_transf_2; 1.
KW Plasmid; Transferase.
SQ SEQUENCE 242 AA; 27360 MW; 503A1E863E48B90C CRC64;

Query Match 6.4%; Score 38; DB 2; Length 242;
Best Local Similarity 34.6%; Pred. No. 4.8;
Matches 9; Conservative 1; Mismatches 16; Indels 0; Gaps

QY 109 LRTEHVVXXLPDXHXXLXXDXDXVYL 134
      | | : ||| | | | | |
Db 179 LAEWAIPQLPDEHATLLNKAREAYL 204

RESULT 10
ID Q49932 PRELIMINARY; PRT; 1446 AA.
AC Q49932;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative polyketide synthase PKSC (PKS).
DS PKSC OR L518_F1.2.
OS Mycobacterium leprae.
SC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson K., Smith D.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE
CC SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -!- COFACTOR: CONTAINS THREE COVALENTLY BOUND PHOSPHOPANTHETINES
CC (POTENTIAL).
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
CC BINDING OF AMP TO THEIR SUBSTRATE.
DR EMBL: U00023; AAA17356.1; -.

```

DR InterPro: IPR001122; Ac-transferase.
DR InterPro: IPR000794; ketoacyl-synt.
DR Pfam: PF00698; Acyl-transf, 1.
DR Pfam: PF00109; ketoacyl-synt, 1.
DR Pfam: PF02801; ketoacyl-synt_C, 1.
DR PROSITE: PS00608; B_KETOACYL-SYNTHASE, 1.
DR Hypothetical protein; Transferase; Acyltransferase;
KW Antibiotic biosynthesis; NADP; Phosphoantethione;

FT	ACT_SITE	204	BETA-RETINOIC ACID SYNTHASE (BY SIMILARITY).
FT	ACT_SITE	651	ACYL TRANSFERASES (BY SIMILARITY).
SQ	SEQUENCE	1446 AA; 153871 MW; 5D23A2E566155FAE CRC64;	
Query Match			
	Best Local Similarity	36.4%;	Score 38; DB 2; Length 1446;
Matches	8; Conservative	1; Mismatches	13; Indels 0; Gaps
QY	112	EHVXXLPDXHXXLXXXXXXYY	133
	:		
Db	990	DHDVSTLHDLNALRGVDNVLY	1011
RESULT 11			
Q9S384		PRELIMINARY;	PRT; 1540 AA.
ID	Q9S384		
AC	Q9S384;		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)	
DT	01-MAR-2002	(TrEMBLrel. 20, Last annotation update)	
DE	Putative polyketide synthase.		
GN	ML2356 OR MLCB12.01C.		

```

Db    1094 KNISQLRTEHRVYELPDEHPLLAQLKREPPDDPCSYLLAIWTP-GETADSIQPSVSTC-- 1150
      |   ||||| | ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    158 XXXXXXXXXCCXXXCCXXXCCXXXCCXXXCCXXXRRGTXL 191
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db    1151 -IFQANGMLCDEETCFSCNSIKETRQIVRGITL 1183
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
RESULT 5
O49498
ID    O49498          PRELIMINARY;             PRT;           917 AA.
AC    O49498;
DT    01-JUN-1998 (TrEMBLrel. 06, Created)
DT    01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    Hypothetical 106.3 kDa protein.
GN    F28A23.180 OR AT4G34060.
OS    Arabidopsis thaliana (Mouse-ear cress).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC    eurosid II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
RA    Herzel A., Neumann S., Hohseisel J., Jesse T., Heijnen L., Vos P.,
RA    Mewes H.W., Mayer K., Schueller C.;
RL    Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP    SEQUENCE FROM N.A.
RA    Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA    Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL    Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP    SEQUENCE FROM N.A.
RA    EU Arabidopsis sequencing project;
RL    Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR    EMBL; AL021961; CAA17566.1; -
DR    EMBL; AL161584; CAB80123.1; -
DR    InterPro; IPR003265; Endo_3c.
DR    InterPro; IPR003651; Fes_bind.
DR    Pfam; PF00730; Hhh-GPD; 1.
DR    SMART; SM00478; ENDO3c; 1.
DR    SMART; SM00525; FES; 1.
KW    Hypothetical protein.
SQ    SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match      13.1%; Score 77; DB 10; Length 917;
Best Local Similarity 44.2%; Pred. No. 2.5e-08;
Matches 19; Conservative 0; Mismatches 18; Indels 6; Gaps 1;

Qy    104 KXXXLRLTEHVXVXLDPDHHXL-----XXDXXXXYLXIXWP 140
      |   ||||| | ||| | | | | | | | | | | | | | | | | |
Db    836 KYNRLRTEHVYVLPDNHELLDFERKKLDSPSYLLAIWQP 878
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

RESULT 6
Q8VWU6
ID    Q8VWU6          PRELIMINARY;             PRT;           53 AA.
AC    Q8VWU6;
DT    01-MAR-2002 (TrEMBLrel. 20, Created)
DT    01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT    01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE    Hypothetical 6.1 kDa protein (Fragment).
OS    Narcissus pseudonarcissus (Daffodil).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
OC    Narcissus.
NCBI_TaxID=39639;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    STRAIN=CV. DUTCH MASTER; TISSUE=TEPAL;
RA    Hunter D.A., Reid M.S.;
```

```

Db      95 - GKGASGKMCFEETCSGNLSREANSQTVRGTTLLIPCRATAMRGSFPLNGTTFVQNELPAD   153
Qy      218 HXXXXXPIXXXXXSLXRXXXXXXXXGXGXXXXXXIXLXLXXXXXXFXGXCRCRXFFXXXXX    277
            |||               |         |         |         |         |         |
Db      154 HessLKPIDPRWDLPDRLPRTVVFGTSVFSGITFRGLSTEQIQFCFWKGFVCVRGEQKTR     213
            |         |         |         |         |         |         |
Qy      278 XPXXLHXXHXHSK          292
            |         |         |         |         |         |         |
Db      214 APRPLYARLHPASK           228

RESULT 2
Q9FTQ2 PRELIMINARY; PRT; 274 AA.
ID Q9FTQ2 AC Q9FTQ2
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE P0665Dl0.14 protein.
GN P0665Dl0.14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC-Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartioideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone:P0665Dl0.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP002861; BABI6489.1; -.
SQ SEQUENCE 274 AA; 30951 MW; D7C836BD3982592D CRC64;

Query Match                24.6%; Score 145; DB 10; Length 274;
Best Local Similarity      26.2%; Pred. No. 2e+24;
Matches 51; Conservative 0; Mismatches 134; Indels 10; Gaps 3;

Qy      104 KXXXXLTREHXVXPDPKHXXLXX-----DXXXXXYLXIWKPPXXXXXXXXXXXXCXKX    157
            |||||       |||        |||       |||       |||       |||       |||
Db      67 KNVSRLTEHQVELPSHLLEGFNQREPDDPCPYLLSIWTP-GETAQSTDAPKSVC---   123

Qy      158 XXXXXXXXCCXXCCXXCCXXCCXXCCXXCGTXLXXXXXXXXXXXXXXXXXXXXXXXFXAD    217
            |||       |||       |||       |||       |||       |||       |||
Db      124 -NSQENGELCASNCFSCNSIREAQAKVKRGTLTLLIPCRTAMRGSFPLNGTYFVNVEVPAD   182

Qy      218 HXXXXXPIXXXXXSLXRXXXXXXXXGXGXXXXXXIXLXLXXXXXXFXGXCRCRXFFXXXXX    277
            |||               |         |         |         |         |         |
Db      183 HDSSRNPDVPRSWINWLPRPTVVFGTSIPTIFKGLTTEIQHCFCWRGVFCVRGFDRTSR    242
            |         |         |         |         |         |         |
Qy      278 XPXXLHXXHXHSK          292
            |         |         |         |         |         |         |
Db      243 APRPLYARLHPASK           257

RESULT 3
Q9SR66 PRELIMINARY; PRT; 1309 AA.
ID Q9SR66 AC Q9SR66
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE T23K18.18 protein.
GN T23K18.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidoisps.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. COLUMBIA;

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 36.2614 Seconds
(without alignments)
1659.222 Million cell updates/sec

Title: US-09-840-743-73

Perfect score:

Sequence: 1 PXXEXPPXXPPXXEXXXXXX.....XXXXXXPPXXLXXXXLHXXXXSK 292

Scoring table: BLOSUM62

Learning rate: 0.00002
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs. 206047115

Total number of hits satisfying chosen parameters:

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 200000000

maximum DE sch length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mnc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvms.*
16: sp_archaeolap.*
17: sp_bacteriap.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query 8			DB	ID	Description
	Score	Match	Length			
1	146	24.7	234	10	Q9L267	Q9L267 arabidopsis
2	145	24.6	274	10	Q9FTQ2	Q9FTQ2 cryza sativ
3	108.5	18.4	1309	10	Q9SR66	Q9SR66 arabidopsis
4	91	15.4	1207	10	Q9SQ06	Q9SJg6 arabidopsis
5	77	13.1	917	10	Q94948	Q94948 arabidopsis
6	47	8.0	53	10	Q8VMU6	Q8VMU6 narcissus p
7	41	6.9	1064	5	Q95U28	Q95U28 drosophila
8	41	6.9	1548	5	Q9VVT7	Q9VVT7 drosophila
9	38	6.4	242	2	Q93CK3	Q93CK3 enterococc
10	38	6.4	1446	2	Q49932	Q49932 mycobacteri
11	38	6.4	1540	16	Q9G384	Q9G384 mycobacteri
12	37	6.3	214	16	Q92SK4	Q92SK4 rhizobium m
13	37	6.3	325	5	Q966M0	Q966M0 caenorhabdi
14	37	6.3	624	5	Q23359	Q23359 caenorhabdi
15	37	6.3	3503	5	Q24292	Q24292 drosophila
16	36	6.1	239	17	Q29807	Q29807 archaeqlob

THIS PAGE BLANK (USPTO)

XX	Example 2; SEQ ID NO 6504; 10078pp; English.
PS	The invention relates to human nucleic acids (AAI57798-AAI61369) and
xx	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
xx	
SQ	Sequence 279 AA;
	Query Match 15.3%; Score 42; DB 22; Length 279;
	Best Local Similarity 41.2%; Pred. No. 32;
	Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY	49 QGXRFXFXWKGVVDSV 65 : : ! : !
Db	81 EGDPEPTQWKGTVLDQV 97

Search completed: November 8, 2002, 19:25:54
Job time : 16.9706 secs

CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 232 AA;
 Query Match 15.3%; Score 42; DB 22; Length 232;
 Best Local Similarity 41.2%; Pred. No. 25;
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 49 QGXRFXXXWKGVSVDV 65
 :| |||:|:|
 Db 34 EGDEPITQWKGTVLDQV 50

RESULT 14
 AAM39788
 ID -AAM39788 standard; Protein; 258 AA.

XX AC AAM39788;
 XX DT 22-OCT-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 2933.

XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PR 09-JUL-2000; 2000US-0598042.
 XX PR 19-JUL-2000; 2000US-0620312.
 XX PR 03-AUG-2000; 2000US-0653450.
 XX PR 14-SEP-2000; 2000US-0662191.
 XX PR 19-OCT-2000; 2000US-0693036.
 XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.
 XX DR N-PSDB; AAI58944.

XX Novel nucleic acids and polypeptides, useful for treating disorders
 XX such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2933; 10078pp; English.

XX-

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nototropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 258 AA;

Query Match 15.3%; Score 42; DB 22; Length 258;
 Best Local Similarity 41.2%; Pred. No. 29;
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 49 QGXRFXXXWKGVSVDV 65
 :| |||:|:|
 Db 60 EGDEPITQWKGTVLDQV 76

RESULT 15
 AAM41573
 ID -AAM41573 standard; Protein; 279 AA.

XX AC AAM41573;

XX DT 22-OCT-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 6504.

XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PR 09-JUL-2000; 2000US-0598042.
 XX PR 19-JUL-2000; 2000US-0620312.
 XX PR 03-AUG-2000; 2000US-0653450.
 XX PR 14-SEP-2000; 2000US-0662191.
 XX PR 19-OCT-2000; 2000US-0693036.
 XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.
 XX DR N-PSDB; AAI60729.

XX Novel nucleic acids and polypeptides, useful for treating disorders
 XX such as central nervous system injuries -

PS Example 1; SEQ ID NO 3451; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 436 AA;

Query Match 15.6%; Score 43; DB 22; Length 436;

Best Local Similarity 32.4%; Pred. No. 36;

Matches 11; Conservative 4; Mismatches 15; Indels 4; Gaps 1;

QY 57 WKGSVDSVXGVFLTQNDXSSXAMXXAXFP 90
| | : : | : | | : | : | : | : |

DB 337 WKACLGVLCLFLTANDGTMTAS----YLAESFP 366

RESULT 12

RAY95051

ID AAU95051 standard; Protein; 466 AA.

XX AC

XX AAU95051;

XX 23-JUN-2000 (first entry)

XX Candida albicans polypeptide sequence # 19.

DE Candida albicans infection; growth; survival; medicament; AIDS;

XX KW vulvovaginitis; immunocompromised patient; treat.

XX OS Candida albicans.

XX EP982401-A2.

XX 01-MAR-2000.

XX 23-DEC-1998; 98EP-0310694.

XX 14-AUG-1998; 98GB-0017796.

XX (JANC) JANSSEN PHARM NV.

XX Contreras RH, Nelissen B, De Backer MD, Luyten WHML, Vlaene JB; Logghe MG;

XX WPI; 2000-258614/23.

XX Essential polypeptides isolated from Candida albicans, useful in the treatment of diseases caused by C.albicans, especially in immunocompromised subjects, e.g., AIDS patients -

XX Claim 3; Page 59-60; 133pp; English.

XX This sequence represents a polypeptide that is critical for the survival and growth of Candida albicans. The C. albicans nucleic acid molecules encoding the polypeptides of the invention may be used as probes and

CC primers for detecting homologous nucleic acid molecule sequences. The polypeptides and nucleic acid molecules and compounds identified as selectively modulating the expression of the polypeptides, may be used as medicaments or for the preparation of a medicament to treat C.albicans associated diseases, especially in AIDS patients and to treat vulvovaginitis in otherwise healthy females. The use of the polypeptides and polynucleotide sequences to treat C.albicans associated diseases has fewer side effects and less toxicity than previously used methods such as the use of amphotericin. This method is therefore especially suitable for immunocompromised patients, such as AIDS patients.

XX SQ Sequence 466 AA;

Query Match 15.6%; Score 43; DB 21; Length 466;

Best Local Similarity 50.0%; Pred. No. 39;

Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 57 WKGSVDSVXGV 68
| : | | | : | :

DB 79 WEGSVLDPIEG 90

RESULT 13

AAM39787

ID AAM39787 standard; Protein; 232 AA.

XX AC AAM39787;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 2932.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

XX OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 23-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AA158943.

XX Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

XX Example 4; SEQ ID NO 2932; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AA42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful

XX WPI; 2001-616774/71.
 DR N-P5DB; AAS59534.
 XX
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 XX Example 1; SEQ ID NO 7822; 1069pp; English.
 PS
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 235 AA;
 SQ

Query Match 17.5%; Score 48; DB 22; Length 235;
 Best Local Similarity 32.0%; Pred. No. 2;
 Matches 8; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 49 QCXRFXXWKGVSVDVXGVFLTON 73
 : : | | | | : :
 Db 60 EGAKTFVSMGSIPLVLLMLIAMN 84
 : : | | | | : :
 RESULT 8
 AAB93159
 ID AAB93159 standard; Protein; 882 AA.
 XX
 AC AAB93159;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:12081.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 XX 07-FEB-2001.
 PD
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX

DR WPI; 2001-318749/34.
 XX
 XX primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 12081; 2537pp + CD ROM; English.
 PS
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 882 AA;
 SQ

Query Match 16.7%; Score 46; DB 22; Length 882;
 Best Local Similarity 33.3%; Pred. No. 24;
 Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 52 RXFXWKGVSVDVXGVFLTONXDXS 78
 : : | | | | : :
 Db 126 RVLTFFPGNLVTSVAGIYFTDNKGDT 152
 : : | | | | : :
 RESULT 9
 AA004564
 ID AA004564 standard; Protein; 106 AA.
 XX
 AC AA004564;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 18456.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 XX 07-SEP-2001.
 PD
 XX 26-FEB-2001; 2001WO-US04927.
 PF
 XX 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX

CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents a DMT-like protein.
XX
SQ Sequence 1729 AA;
Query Match 59.6%; Score 164; DB 23; Length 1729;
Best Local Similarity 43.0%; Pred. No. 1.3e-20;
Matches 43; Conservative 0; Mismatches 47; Indels 10; Gaps 2;
QY 1 KVVXDXXTXXWXXLXXXXXX-----XXERXXFXFXFXFXFXMXXXXQGX 51
DB 697 KVDIDETTRINWLMGKGEKDEKDKKEKWEERVRFRGRADSFIAARMHLVQGD 756
QY 52 RFXPKGWSVDSVGVFLTON-XDXXSSXAXMXXAXXFP 90
DB 757 RFXPKGWSVDSVGVFLTONVSDHLSSAFMSLAARFP 796
RESULT 4
AAU72737
ID AAU72737 standard; Protein; 1413 AA.
XX
AC AAU72737;
XX
DT 26-FEB-2002 (first entry)
XX
DE Arabidopsis DMT2 (1DMT2) protein.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Arabidopsis Thaliana.
XX
PN WO200180626-A1.
XX
PD 01-NOV-2001.
XX
XX 23-APR-2001; 2001WO-US13059.
XX
XX 21-APR-2000; 2000US-0553690.
XX
XX (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
XX WPI; 2002-055307/07.
DR N-PSDB; AAS96692.
XX
XX New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
XX
XX Disclosure; Page 67; 109pp; English.
XX
XX The invention relates to an isolated polynucleotide sequence or their
XX complement encoding a polypeptide having a sequence at least 40%
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
XX or C or their combinations. Also included are an expression

CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents a DMT-like protein.
XX
SQ Sequence 1413 AA;
Query Match 59.5%; Score 163.5; DB 23; Length 1413;
Best Local Similarity 43.4%; Pred. No. 1.2e-20;
Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;
QY 1 KVVXDXXTXXWXXL-----XXXDXDXDXDXDXDXDXFXFXMXXXXQGX 52
DB 528 KVDLDETDTRVWKLLENINSEGVDSDEQAKAKWEEERNVFRGRADSFIAARMHLVQGD 587
QY 53 RFXPKGWSVDSVGVFLTON-XDXXSSXAXMXXAXXFP 90
DB 588 RFXPKGWSVDSVGVFLTONVSDHLSSAFMSLASQFP 626
RESULT 5
AAU72743
ID AAU72743 standard; Protein; 1952 AA.
XX
AC AAU72743;
XX
DT 26-FEB-2002 (first entry)
XX
DE Rice DMT1 protein.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Oryza sativa.
XX
PN WO200180626-A1.
XX
PD 01-NOV-2001.
XX
XX 23-APR-2001; 2001WO-US13059.
XX
XX 21-APR-2000; 2000US-0553690.
XX
XX (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
XX WPI; 2002-055307/07.
DR N-PSDB; AAS96695.
XX
XX New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
XX
XX Disclosure; Page 79-80; 109pp; English.
XX
XX The invention relates to an isolated polynucleotide sequence or their

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:08 ; Search time 13.9706 Seconds
(without alignments)
858.415 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KVXXDXXTXWXXLXXXXD.....TQNXDXSSXAMXAXXFP 90

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166.5	60.5	1309	23 ABB92260	Herbicideally activ
2	166.5	60.5	1332	23 AAU72739	Arabidopsis DMT3 (
3	164	59.6	1729	23 AAU72736	Arabidopsis Demete
4	163.5	59.5	1413	23 AAU72737	Arabidopsis DMT2 (
5	162.5	59.1	1952	23 AAU72743	Rice DMT1 protein.
6	162	58.9	1114	23 AAU72741	Arabidopsis DMT4 (
7	48	17.5	235	22 AAU46627	Propionibacterium
8	46	16.7	882	22 AAB93159	Human protein sequ
9	44	16.0	106	22 AAO04564	Human polypeptide
10	43	15.6	434	23 ABB48287	Listeria monocytog

11	43	15.6	436	22 AAU42256	Propionibacterium
12	43	15.6	466	21 AAY95051	Candida albicans p
13	42	15.3	232	22 AAM39787	Human polypeptide
14	42	15.3	258	22 AAM39788	Human polypeptide
15	42	15.3	279	22 AAM41573	Human polypeptide
16	42	15.3	279	22 AAM41574	Human polypeptide
17	42	15.3	370	22 AAB93450	Human protein sequ
18	42	15.3	370	23 AAB47993	Pristanoyl-CoA oxi
19	42	15.3	973	19 AAU48304	Amino acid sequenc
20	41.5	15.1	351	22 AAU50778	Propionibacterium
21	41	14.9	331	22 ABG10353	Novel human diagno
22	41	14.9	453	22 ABG12472	Novel human diagno
23	41	14.9	973	23 ABB97191	Novel human protei
24	41	14.9	973	23 ABB05701	Human intracellula
25	40.5	14.7	84	20 AAY11870	Human 5' EST secre
26	40.5	14.7	310	23 ABB92345	Herbicideally activ
27	40	14.5	159	23 ABB38906	Staphylococcus epi
28	40	14.5	196	23 ABB30378	Streptococcus poly
29	40	14.5	200	23 ABB28344	Streptococcus poly
30	40	14.5	211	21 AAY51682	Murine clodin 7 pr
31	40	14.5	313	21 AAY90872	Human G protein-co
32	40	14.5	313	22 AAU10306	G-protein coupled
33	40	14.5	313	22 AAG72077	Human olfactory re
34	40	14.5	318	20 AAW90108	Human FLAME-2 prot
35	40	14.5	318	20 AAW90109	Mouse FLAME-2 prot
36	40	14.5	318	21 AAY51022	Human DEDD protein
37	40	14.5	318	21 AAY51023	Murine DEDD protei
38	40	14.5	318	22 AAB93016	Human protein sequ
39	40	14.5	318	22 AAB94040	Human protein sequ
40	40	14.5	1140	22 AAE09365	Human ATP-binding
41	40	14.5	1182	22 AAM93387	Human polypeptide,
42	40	14.5	1503	20 AAY43544	A human MPR-relate
43	40	14.5	1503	22 AAE09361	Human ATP-binding
44	40	14.5	1503	22 AAE09363	Human ATP-binding
45	40	14.5	1503	22 AAE09364	Human ATP-binding

ALIGNMENTS

RESULT 1
ABB92260
ID ABB92260 standard; Protein; 1309 AA.
XX
AC ABB92260;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicideally active polypeptide SEQ ID NO 1471.
XX
KW Herbicideal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
FA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicideally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX

CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,236
FILING DATE: March 11, 1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: EXEL98-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 634 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-236-2

Query Match 13.8%; Score 38; DB 4; Length 634;
Best Local Similarity 58.3%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 50 GRRFXFXWKGSV 61
| | | | |
Db 19 GPRIFAVWKGHV 30

RESULT 15

US-09-240-410-2
Sequence 2, Application US/09240410
Patent No. 6197544
GENERAL INFORMATION:
APPLICANT: MICHALOVICH, DAVID
APPLICANT: HAYES, PHILIP DAVID
TITLE OF INVENTION: NOVEL COMPOUNDS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/240,410
FILING DATE: 27-JAN-1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK APPLICATION NO. TO BE ASSIGNED
FILING DATE: 20-JAN-1999
APPLICATION NUMBER: EP APPLICATION NO. 98300694.1
FILING DATE: 30-JAN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-30039
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-240-410-2

Query Match 13.8%; Score 38; DB 4; Length 666;
Best Local Similarity 58.3%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 50 GRRFXFXWKGSV 61
| | | | |
Db 51 GPRIFAVWKGHV 62

Search completed: November 8, 2002, 19:29:14
Job time : 7.14706 secs

GENERAL INFORMATION:
APPLICANT: Garner, Ian
APPLICANT: Dalrymple, Michael A
APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of Fibrinogen in Transgenic
TITLE OF INVENTION: Animals
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206.176
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 93-15
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 491 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-206-176-4

Query Match 14.2%; Score 39; DB 1; Length 491;
Best Local Similarity 47.1%; Pred. No. 78;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 57 WKGSVVDVXGVFLTON 73
Db 347 WKGDVKAHYGGFTVQN 363

RESULT 12
US-09-620-412C-317
Sequence 317, Application US/09620412C
Patent No. 6448234
GENERAL INFORMATION:
APPLICANT: Steven P. Fling
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
FILE REFERENCE: 210121.469C7
CURRENT APPLICATION NUMBER: US/09/620.412C
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 317
LENGTH: 646
TYPE: PRT
ORGANISM: Chlamydia trachomatis
US-09-620-412C-317

Query Match 14.2%; Score 39; DB 4; Length 646;
Best Local Similarity 27.0%; Pred. No. 1.1e+02;
Matches 10; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

Qy 37 RXXFXIXMXXQGXRXFXWKGSVVDVXGVFLTON 73

Db 326 RTLAFINNMSGDCGSAISADTQISITDVKILFENN 362

RESULT 13
US-08-602-359A-33
Sequence 33, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:

APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFFIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602.359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-602-359A-33

Query Match 13.8%; Score 38; DB 2; Length 184;
Best Local Similarity 45.8%; Pred. No. 37;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 2;

Qy 50 GXRFXFXW---KGSVVDVXGVFL 70
Db 72 GLRNLYSWSRVKGSLLISM-GVFL 94

RESULT 14
US-09-041-236-2
Sequence 2, Application US/09041236
Patent No. 6225285
GENERAL INFORMATION:
APPLICANT: Luo, Yuling
APPLICANT: Xiomel, Xu
TITLE OF INVENTION: Semaphorin K1
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE

```
RESULT 8
US-09-276-993-4
; Sequence 4, Application US/09276993
; Patent No. 6207801
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Enad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6207801Iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-4
;
; Query Match 14.5%; Score 40; DB 4; Length 318;
; Best Local Similarity 33.3%; Pred. No. 31;
; Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVDSVXGVFTQN 73
I: ||::: |||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 9
US-09-276-993-6
; Sequence 6, Application US/09276993
; Patent No. 6207801
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Enad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6207801Iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-6
;
; Query Match 14.5%; Score 40; DB 4; Length 318;
; Best Local Similarity 33.3%; Pred. No. 31;
; Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVDSVXGVFTQN 73
I: ||::: |||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 10
US-09-213-053-3
; Sequence 3, Application US/09213053
; Patent No. 6159477
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe
; APPLICANT: BAUDU, Philippe
; TITLE OF INVENTION: CANINE HERPESVIRUS BASED RECOMBINANT LIVE VACCINE, IN
; TITLE OF INVENTION: PARTICULAR AGAINST CANINE DISTEMPER, RABIES OR THE
; FILE OF INVENTION: PARAINFLUENZA 2 VIRUS
; FILE REFERENCE: 454313-2200
; CURRENT APPLICATION NUMBER: US/09/213,053
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 9608242
; EARLIER FILING DATE: 1996-06-27
; EARLIER APPLICATION NUMBER: PCT/FR97/01115
; EARLIER FILING DATE: 1997-06-23
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Canine herpesvirus
; US-09-213-053-3
;
; Query Match 14.2%; Score 39; DB 4; Length 398;
; Best Local Similarity 36.8%; Pred. No. 61;
; Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 52 RFXFXWKSGVDSVXGVFL 70
I: |::|::|::|
Db 22 RIWKTWRNSIKDNVCGEIL 40

RESULT 11
US-08-206-176-4
; Sequence 4, Application US/08206176
; Patent No. 5639940
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-6
;
; Query Match 14.5%; Score 40; DB 4; Length 318;
; Best Local Similarity 33.3%; Pred. No. 31;
; Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVDSVXGVFTQN 73
I: ||::: |||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 10
US-09-213-053-3
; Sequence 3, Application US/09213053
; Patent No. 6159477
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe
; APPLICANT: BAUDU, Philippe
; TITLE OF INVENTION: CANINE HERPESVIRUS BASED RECOMBINANT LIVE VACCINE, IN
; TITLE OF INVENTION: PARTICULAR AGAINST CANINE DISTEMPER, RABIES OR THE
; FILE OF INVENTION: PARAINFLUENZA 2 VIRUS
; FILE REFERENCE: 454313-2200
; CURRENT APPLICATION NUMBER: US/09/213,053
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 9608242
; EARLIER FILING DATE: 1996-06-27
; EARLIER APPLICATION NUMBER: PCT/FR97/01115
; EARLIER FILING DATE: 1997-06-23
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Canine herpesvirus
; US-09-213-053-3
;
; Query Match 14.2%; Score 39; DB 4; Length 398;
; Best Local Similarity 36.8%; Pred. No. 61;
; Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 52 RFXFXWKSGVDSVXGVFL 70
I: |::|::|::|
Db 22 RIWKTWRNSIKDNVCGEIL 40

RESULT 11
US-08-206-176-4
; Sequence 4, Application US/08206176
; Patent No. 5639940
```

STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/859,167
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-859-167-6

Query Match 14.5%; Score 40; DB 3; Length 318;
Best Local Similarity 33.3%; Pred. No. 31;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVSDSVXGVFLTON 73
I: ||:||||: |||:|
Db 260 WRDYINGSLLLEALKGVFITDS 280

RESULT 6
US-09-109-273-4
Sequence 4, Application US/09109273
Patent No. 6063760
GENERAL INFORMATION:
APPLICANT: Alnemri, Emad S.
APPLICANT: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
TITLE OF INVENTION: OF MAKING THE SAME
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6063760ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/109,273
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/859,167
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-109-273-6

Query Match 14.5%; Score 40; DB 3; Length 318;
Best Local Similarity 33.3%; Pred. No. 31;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVSDSVXGVFLTON 73
I: ||:||||: |||:|
Db 260 WRDYINGSLLLEALKGVFITDS 280

TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-109-273-4

Query Match 14.5%; Score 40; DB 3; Length 318;
Best Local Similarity 33.3%; Pred. No. 31;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVSDSVXGVFLTON 73
I: ||:||||: |||:|
Db 260 WRDYINGSLLLEALKGVFITDS 280

RESULT 7
US-09-109-273-6
Sequence 6, Application US/09109273
Patent No. 6063760
GENERAL INFORMATION:
APPLICANT: Alnemri, Emad S.
APPLICANT: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
TITLE OF INVENTION: OF MAKING THE SAME
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6063760ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/109,273
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/859,167
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-109-273-6

Query Match 14.5%; Score 40; DB 3; Length 318;
Best Local Similarity 33.3%; Pred. No. 31;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVSDSVXGVFLTON 73
I: ||:||||: |||:|
Db 260 WRDYINGSLLLEALKGVFITDS 280

```

US-09-517-639-4
; Sequence 4, Application US/09517639
; Patent No. 6414120
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo M.
; TITLE OF INVENTION: Mammalian Deep Orange Proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,639
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,452
; FILING DATE: 31-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1240.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-517-639-4

```

```

Query Match          14.5%; Score 40; DB 4; Length 159;
Best Local Similarity 24.3%; Pred. No. 14;
Matches          9; Conservative          5; Mismatches          23; Indels          0; Gaps

Qy      54 FXXWKGSVDSVXGVFLTONXDXSSXAXMXXXXP 90
      |  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db      92 FATSGKEGDTSMPGMHITLAFEEVKGKTTVTSTSTP 128

RESULT 4
US-08-859-167-4
; Sequence 4, Application US/08859167
; Patent No. 6037461
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS OF
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6037461iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/859,167
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-859-167-4

Query Match          14.5%; Score 40; DB 3; Length 318;
Best Local Similarity 33.3%; Pred. No. 31;
Matches          7; Conservative          8; Mismatches          2; Indels          4; Gaps

Qy      57 WK----GSVVDXSVXGVFLTON 73
      |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      260 WRDYINGSLLEALKGVFITDS 280

RESULT 5
US-08-859-167-6
; Sequence 6, Application US/08859167
; Patent No. 6037461
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS OF
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6037461iris

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 5.14706 Seconds
(without alignments)
514.481 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KVVXXXTXXXXXLLXXXX.....TQNXDXXSSXAMXXAXXFP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	15.3	973	US-08-904-452-4	Sequence 4, Appli
2	42	15.3	973	US-09-517-639-4	Sequence 4, Appli
3	40	14.5	159	US-09-134-001C-3751	Sequence 3751, Ap
4	40	14.5	318	US-08-859-167-4	Sequence 4, Appli
5	40	14.5	318	US-08-859-167-6	Sequence 6, Appli
6	40	14.5	318	US-09-109-273-4	Sequence 4, Appli
7	40	14.5	318	US-09-109-273-6	Sequence 6, Appli
8	40	14.5	318	US-09-276-993-4	Sequence 4, Appli
9	40	14.5	318	US-09-276-993-6	Sequence 6, Appli
10	39	14.2	398	US-09-213-053-3	Sequence 3, Appli
11	39	14.2	491	US-08-206-176-4	Sequence 4, Appli
12	39	14.2	646	US-09-620-412C-317	Sequence 317, App
13	38	13.8	184	US-08-602-359A-33	Sequence 33, Appli
14	38	13.8	634	US-09-041-236-2	Sequence 2, Appli
15	38	13.8	666	US-09-240-410-2	Sequence 2, Appli
16	37.5	13.6	1198	US-09-199-637A-405	Sequence 405, App
17	37	13.5	138	US-09-057-762-13	Sequence 13, Appli
18	37	13.5	138	US-08-326-119A-13	Sequence 13, Appli
19	37	13.5	271	US-08-433-263B-2	Sequence 2, Appli
20	37	13.5	301	US-09-134-001C-4909	Sequence 4909, Ap
21	37	13.5	1127	US-08-937-195-3	Sequence 3, Appli
22	37	13.5	1127	US-08-915-152-3	Sequence 3, Appli
23	37	13.5	1127	PCT-US96-07627-3	Sequence 3, Appli
24	37	13.5	15281	US-08-471-119A-2	Sequence 2, Appli
25	36.5	13.3	175	US-08-858-207A-363	Sequence 363, App
26	36.5	13.3	245	US-08-715-204-3	Sequence 3, Appli
27	36.5	13.3	245	US-09-162-597-3	Sequence 3, Appli

28	36	13.1	1213	2	US-08-937-102-2	Sequence 2, Appli
29	36	13.1	1443	2	US-08-670-707A-39	Sequence 39, Appli
30	36	13.1	1443	4	US-09-037-601-39	Sequence 39, Appli
31	36	13.1	1443	4	US-09-315-179-39	Sequence 39, Appli
32	36	13.1	1467	4	US-09-523-656-38	Sequence 38, Appli
33	36	13.1	2115	4	US-09-324-867-5	Sequence 5, Appli
34	36	13.1	2133	2	US-08-670-707A-37	Sequence 37, Appli
35	36	13.1	2133	4	US-09-037-601-37	Sequence 37, Appli
36	36	13.1	2133	4	US-09-315-179-37	Sequence 37, Appli
37	36	13.1	2133	4	US-09-523-656-30	Sequence 30, Appli
38	36	13.1	3898	2	US-08-876-991-2	Sequence 2, Appli
39	36	13.1	3898	2	US-09-059-853-2	Sequence 2, Appli
40	35.5	12.9	365	4	US-09-151-592-2	Sequence 2, Appli
41	35.5	12.9	365	4	US-09-254-077A-10	Sequence 10, Appli
42	35	12.7	90	4	US-08-936-165A-270	Sequence 270, App
43	35	12.7	108	4	US-09-134-001C-3747	Sequence 3747, Ap
44	35	12.7	161	4	US-09-413-814-15	Sequence 15, Appli
45	35	12.7	184	1	US-08-075-193-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-904-452-4
; Sequence 4, Application US/08904452
; Patent No. 6083742
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo M.
; TITLE OF INVENTION: Mammalian Deep Orange Proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904.452
; FILING DATE: 31-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1240.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-904-452-4

Query Match 15.3%; Score 42; DB 3; Length 973;
Best Local Similarity 47.1%; Pred. No. 51;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 49 QGXRFXWKGVSVDV 65
Db 129 OKARPLAWKGVESV 145

RESULT 2

THIS PAGE BLANK (USPTO)

```
QY 67 GVFLTQNDXSSXAMXXA 86
      | : | | | | | | | |
Db 86 GIFTTPEDETSSATTISTS 105

RESULT 13
US-09-836-077-3
; Sequence 3, Application US/09836077
; Patent No. US20020037851A1
; GENERAL INFORMATION:
; APPLICANT: FLECKENSTEIN, Bernhard
; TITLE OF INVENTION: HUMAN SEMAPHORIN L (H-SEMAL) AND
; CORRESPONDING SEMAPHORINS IN OTHER SPECIES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frommer Lawrence & Haug LLP
; STREET: 745 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/836,077
; FILING DATE: 16-Apr-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence, William F.
; REGISTRATION NUMBER: 28,029
; REFERENCE/DOCKET NUMBER: 514429-3647
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 3:
; *
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; *
; MOLECULE TYPE: amino acid
; *
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-836-077-3
Query Match 13.8%; Score 38; DB 10; Length 666;
Best Local Similarity 58.3%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 50 GXRFXFXWKGSV 61
      | : | | | | | | | |
Db 51 GPRIFAVWKGHV 62

RESULT 14
US-09-764-587A-2
; Sequence 2, Application US/09764587A
; Patent No. US20020106722A1
; GENERAL INFORMATION:
; APPLICANT: David Michalovich
; APPLICANT: Philip David Hayes
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30039-D1
; CURRENT APPLICATION NUMBER: US/09/764,587A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 09/240,410
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: EP 98300694.1
; PRIOR FILING DATE: 1999-01-30
```

```
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 666
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-764-587A-2

Query Match 13.8%; Score 38; DB 10; Length 666;
Best Local Similarity 58.3%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 50 GXRFXFXWKGSV 61
      | : | | | | | | | |
Db 51 GPRIFAVWKGHV 62

RESULT 15
US-09-799-875-2
; Sequence 2, Application US/09799875
; Patent No. US20020034780A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: No. US20020034780A1 Human Protein Kinases and Uses
; FILE REFERENCE: 35800/209996
; CURRENT APPLICATION NUMBER: US/09/799,875
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/182,059
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/659,287
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-799-875-2

Query Match 13.8%; Score 38; DB 10; Length 879;
Best Local Similarity 30.3%; Pred. No. 72;
Matches 10; Conservative 2; Mismatches 21; Indels 0; Gaps 0;

QY 58 KGSVVDSVXGVFLTQNDXSSXAMXXAXFP 90
      | : | | | | | | | |
Db 542 KNSKRDYLSSTFLCSDDDDRASKNISMNSDSSFP 574

Search completed: November 8, 2002, 19:29:40
Job time : 4.79412 secs
```

```
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-317

Query Match 14.2%; Score 39; DB 10; Length 646;
Best Local Similarity 27.0%; Pred. No. 34;
Matches 10; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

Qy 37 RXXFXFXMXXXXGXRXFXFXKXGVSXVGVFELTON 73
      | | | | | | | | | | | | | | | | | |
Db 326 RFLAFINMNSGDCGGAISADTQISTITVKGILFENN 362

RESULT 10
US-09-903-410-33
; Sequence 33, Application US/09903410
; Patent No. US20020146799A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: MURPHY, Dennis
; APPLICANT: REID, John
; APPLICANT: MAFFIA, Anthony
; APPLICANT: LINK, Steven
; APPLICANT: SWANSON, Ronald
; APPLICANT: WARREN, Patrick
; APPLICANT: KOSMOTKA, Anna
; TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: DIVER1180-2
; CURRENT APPLICATION NUMBER: US/09/903,410
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 09/382,242
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/602,359
; PRIOR FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Staphylothermus Marinus
US-09-903-410-33

Query Match 13.8%; Score 38; DB 10; Length 184;
Best Local Similarity 45.8%; Pred. No. 12;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 2;

Qy 50 GXRFXFXW---KGSVDSVXGVFL 70
      | | | | | | | | | | | | | | | | | |
Db 72 GLRNLWSVRKGLISM-GVFL 94

RESULT 11
US-09-935-390A-36
; Sequence 36, Application US/09935390A
; Patent No. US200200761A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; APPLICANT: Quianjín, Hu
; APPLICANT: Garcia, Pablo
; APPLICANT: Williams, Lewis T.
; APPLICANT: Kothakota, Srinivas
; TITLE OF INVENTION: Secreted Human Proteins
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/935,390A
FILING DATE: 22-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/988,671
FILING DATE: 1997-12-11
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. R. Potter
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 1369.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2718
TELEFAX: (510) 655-3542
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. US20020076761a1
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-935-390A-36

Query Match 13.8%; Score 38; DB 10; Length 210;
Best Local Similarity 24.4%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 26; Indels 0; Gaps 0;

Qy 50 GXRFXFXKXGVSXVGVFELTONXXSSXAXMXXAXFP 90
      | | | | | | | | | | | | | | | | | |
Db 162 GPAFIGWAGSALVILGALLSCSCPGNESKAGYAPRSYP 202

RESULT 12
US-09-764-864-1235
; Sequence 1235, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1235
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (488)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (493)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (494)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1235

Query Match 13.8%; Score 38; DB 10; Length 494;
Best Local Similarity 35.0%; Pred. No. 37;
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
```

```

: PRIOR APPLICATION NUMBER: 60/178,227
: PRIOR FILING DATE: 2000-01-26
: PRIOR APPLICATION NUMBER: 60/220,590
: PRIOR FILING DATE: 2000-07-25
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 86
: LENGTH: 305
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-761-288-86

Query Match 14.2% Score 39; DB 10; Length 305;
Best Local Similarity 38.9%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 54 FXXKSGVSDSVXGVFLT 71
Db 146 FALWLGTHSLVQTLT 163
+ + + + +
: : : : :

RESULT 8
US-09-825-414-56
: Sequence 56, Application US/09825414
: Patent No. US20020083489A1
: GENERAL INFORMATION:
: APPLICANT: Collmer, Alan
: APPLICANT: Alfano, James R.
: APPLICANT: Charkowski, Amy O.
: TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
: TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
: FILE REFERENCE: 19603/3243
: CURRENT APPLICATION NUMBER: US/09/825,414
: CURRENT FILING DATE: 2001-04-03
: PRIOR APPLICATION NUMBER: 60/194,160
: PRIOR FILING DATE: 2000-04-03
: PRIOR APPLICATION NUMBER: 60/224,604
: PRIOR FILING DATE: 2000-08-11
: PRIOR APPLICATION NUMBER: 60/249,548
: PRIOR FILING DATE: 2000-11-17
: NUMBER OF SEQ ID NOS: 91
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 56
: LENGTH: 316
: TYPE: PRT
: ORGANISM: Pseudomonas syringae pv. delphinii
US-09-825-414-56

Query Match 14.2% Score 39; DB 10; Length 316;
Best Local Similarity 53.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 58 KGSVSDSVXGVFL 70
Db 101 KGLIDLVSAGFL 113
+ + + + +
: : : : :

RESULT 9
US-09-841-132-317
: Sequence 317, Application US/09841132
: Patent No. US20020061848A1
: GENERAL INFORMATION:
: APPLICANT: Bhatia, Ajay
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Probst, Peter
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
: FILE REFERENCE: 210121.469C8
: CURRENT APPLICATION NUMBER: US/09/841,132
: CURRENT FILING DATE: 2001-04-23
: NUMBER OF SEQ ID NOS: 599
: SOFTWARE: FastSeq for Windows Version 3.0/4.0
: SEQ ID NO 317

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:26:04 ; Search time 2.79412 Seconds
(without alignments)
464.428 Million cell updates/sec

Title: US-09-840-743-71
 Perfect score: 275
 Sequence: 1 KVYDXXTXXWXXLXXDXD.....TONYDXXSSXAMXXAXXEP 90

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

Maximum DB seq length:	200000000
Maximum DB seq length:	200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pcp.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pcp.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pcp.*
10: /cgn2_6/ptodata/1/pubaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pcp.*
12: /cgn2_6/ptodata/1/pubaa/US10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pcp.*
14: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	40	14.5	210	10	US-09-733-167-6		Sequence 6, Appli
2	40	14.5	318	10	US-09-733-167-1		Sequence 1, Appli
3	40	14.5	318	10	US-09-733-167-3		Sequence 3, Appli
4	39.5	14.4	164	10	US-09-815-242-13509		Sequence 13630, A
5	39.5	14.4	164	10	US-09-815-242-13630		Sequence 13509, A
6	39	14.2	241	10	US-09-825-414-22		Sequence 22, Appl
7	39	14.2	305	10	US-09-761-288-86		Sequence 86, Appl
8	39	14.2	316	10	US-09-825-414-56		Sequence 56, Appl
9	39	14.2	646	10	US-09-841-132-317		Sequence 317, App
10	38	13.8	184	10	US-09-903-410-33		Sequence 33, Appl
11	38	13.8	210	10	US-09-935-390A-36		Sequence 36, Appl
12	38	13.8	494	10	US-09-764-864-1235		Sequence 1235, Ap
13	38	13.8	666	10	US-09-836-077-3		Sequence 3, Appli
14	38	13.8	666	10	US-09-764-587A-2		Sequence 2, Appli
15	38	13.8	879	10	US-09-799-875-2		Sequence 2, Appli
16	37.5	13.6	167	10	US-09-815-242-13543		Sequence 13543, A
17	37.5	13.6	792	10	US-09-995-587A-11		Sequence 11, Appl
18	37	13.5	217	10	US-09-799-777-26		Sequence 26, Appl
19	37	13.5	276	10	US-09-815-242-11729		Sequence 11729, A

20	37	13.5	453	10	US-09-864-761-37331
21	36.5	13.3	138	12	US-10-052-586-6
22	36	13.1	36	10	US-09-864-761-43486
23	36	13.1	53	10	US-09-864-761-42021
24	36	13.1	272	10	US-09-815-242-11074
25	36	13.1	310	10	US-09-886-055-153
26	36	13.1	479	9	US-09-992-598-216
27	36	13.1	479	10	US-09-989-722-216
28	36	13.1	479	10	US-09-989-723-216
29	36	13.1	479	10	US-09-989-727-216
30	36	13.1	479	10	US-09-989-727-216
31	36	13.1	479	10	US-09-989-731-216
32	36	13.1	479	10	US-09-989-732-216
33	36	13.1	479	10	US-09-989-732-216
34	36	13.1	479	10	US-09-91-073-216
35	36	13.1	479	10	US-09-900-442-216
36	36	13.1	479	10	US-09-931-163-216
37	36	13.1	479	10	US-09-993-604-216
38	36	13.1	479	10	US-09-990-456-216
39	35.5	12.9	565	10	US-09-989-721-216
40	35.5	12.9	552	10	US-09-051-034A-4
41	35	12.7	59	10	US-09-815-242-11002
42	35	12.7	90	10	US-09-864-761-41743
43	35	12.7	99	10	US-09-939-880-270
44	35	12.7	118	10	US-09-911-826A-22
45	35	12.7	107	10	US-09-816-248-1
					US-09-815-242-4958

ALIGNMENTS

```

RESULT 1
US-09-733-167-6
; Sequence 6, Application US/09733167
; Patent No. US20020099009A1
; GENERAL INFORMATION:
;   APPLICANT: Peter, Marcus
;   APPLICANT: Krammer, Peter
;   TITLE OF INVENTION: Protein for Regulation of Apoptosis
;   FILE REFERENCE: 4121-120
;   CURRENT APPLICATION NUMBER: US/09/733.167
;   CURRENT FILING DATE: 2000-12-08
;   PRIOR APPLICATION NUMBER: PCT/DE99/01712
;   PRIOR FILING DATE: 1999-06-08
;   PRIOR APPLICATION NUMBER: German Patent Application No.
;   PRIOR FILING DATE: 1998-06-08
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 210
;   TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
;   OTHER INFORMATION: C-DEDD, which is a deletion mutant
;   -OTHER INFORMATION: acids 109-318 of the naturally occur
US-09-733-167-6

```

Qy 57 WK----GSVVDVSXGVFLTN 73
| : | : : : | : :
Db 152 WRDYINGSLLLEALKGVFITDS 172

RESULT 2
US-09-733-167-1
; Sequence 1, Application US/09733167
; Patent No. US2002009909A1
; GENERAL INFORMATION:
; APPLICANT: Peter, Marcus
; APPLICANT: Krammer, Peter

RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=98447599; PubMed=9774341;
RA Stegh A.H., Schickling O., Ehret A., Scaffidi C., Peterhaensel C.,
RA Hofmann T.G., Grummt I., Krammer P.H., Peter M.E.;
RT "DEDD, a novel death effector domain-containing protein, targeted to
RT the nucleolus";
RL EMOB J. 17:5974-5986(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC TISSUE=Testis;
RX MEDLINE=95049260; PubMed=9832420;
RA Leo C.P., Hsu S.Y., McGee E.A., Salanova M., Hsueh A.J.W.;
RT "DEFT, a novel death effector domain-containing molecule predominantly
RT expressed in testicular germ cells";
RL Endocrinology 139:4839-4848(1998).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Pan G.;
RT "FLDED-1, a novel molecule with a DED-like domain.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Thome M., Tschopp J.;
RT "DEDPOL, a novel DED-containing protein.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Blood;
RA Zhao Z., Huang X., Li N., Zhu X., Cao X.;
RT "A novel gene from human dendritic cell.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Bone marrow, and Placenta;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binds strongly and non-specifically to DNA. Induces
CC apoptosis. Inhibits DNA transcription in vitro (By
CC similarity).
CC -!- SUBUNIT: Interacts with caspase-8 and FADD (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Translocated to the nucleus
CC during CD95-mediated apoptosis where it is localized in the
CC nucleoli (By similarity).
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Widely expressed with highest levels in
CC testis.
CC -!- SIMILARITY: CONTAINS 1 DEATH EFFECTOR DOMAIN (DED).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF083236; AAC33105.1; -
DR EMBL; AF100341; AAD16414.1; -
DR EMBL; AF043733; AAC80280.1; -
DR EMBL; AJ010973; CA09445.1; -
DR EMBL; AF064605; AAC17110.3; -
DR EMBL; BC016724; AAH16724.1; -
DR EMBL; BC013910; AAH13910.1; -
DR Genew; HGNC:2755; DEDD.
DR MIM; 606841; -
DR InterPro; IPR001875; DED.
DR Pfam; PF01335; DED; 1.
DR SMART; SM00031; DED; 1.
DR PROSITE; PS0168; DED; 1.
KW Apoptosis; Transcription regulation; Repressor; DNA-binding;
KW Nuclear protein; Alternative splicing.
FF DOMAIN 25 103

FT VARSPLIC 194 194 D-> GEEIQGFQWRSLEGEYKELGHVAVIAIQY (IN
FT ISOFORM 2).
FT CONFLICT 13 13 P-> L (IN REF. 5).
SQ SEQUENCE 318 AA; 36794 MW; FF9D5FF9B61F6BB6 CRC64;
Query Match 14.5%; Score 40; DB 1; Length 318;
Best Local Similarity 33.3%; Pred. No. 24;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;
Qy 57 WK---GSVVDVSVXGVFLTN 73
I: ||::: |||:|
Db 260 WRDYINGSLLEALKGVFITDS 280
Search completed: November 8, 2002, 19:27:09
Job time : 5.23529 secs

```
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE000030; AAB95974.1; -.
DR HSSP; Q9KWU7; 1HQM.
DR InterPro; IPR001572; RNA_pol_B.
DR Pfam; PF00562; RNA_pol_B; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
DR Transferase; Transcription; DNA-directed RNA polymerase;
KW Complete proteome.
SQ SEQUENCE 1391 AA; 155621 MW; B2F345AB24F18EAD CRC64;

Query Match 14.9%; Score 41; DB 1; Length 1391;
Best Local Similarity 64.3%; Pred. No. 78;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 58 KGSVVDVSVXGVFLT 71
| | | | |
Db 118 KNSATSSVDGVFLT 131

RESULT 13
CLD7_RAT
ID CLD7_RAT STANDARD; PRT; 191 AA.
AC Q921L1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 39, Last annotation update)
DE Claudin-7 (Fragment).
GN CLDN7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ011811; CAA09790.1; -.
DR InterPro; IPR001832; Claudin.
DR Pfam; PF00822; PMP22_Claudin; 1.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 1 1
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
SQ SEQUENCE 191 AA; 20366 MW; 83B445908DFF41A CRC64;

Query Match 14.5%; Score 40; DB 1; Length 191;
Best Local Similarity 24.4%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 26; Indels 0; Gaps 0;

Qy 50 GXRFXFXWKSVDVSVXGVFLTQNXDXSSXAXMXXAXXFP 90
| | | | |
Db 142 GPAIFGWAGSALVLLGGLSCPCGSGSKAAYRAPRSYP 182

RESULT 14
CLD7_MOUSE
ID DEDD_HUMAN STANDARD; PRT; 318 AA.
AC O75618; O60737;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Death effector domain-containing protein (death effector domain-
DE containing testicular molecule) (DEDProl) (FLDED-1) (K605).
GN DEDD OR DEFT OR DEDPRO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```


RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RL *subtilis*.";
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE GPF0053 FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y14082; CAA74500.1; -;
DR EMBL: Z99109; CAB12794.1; -;
DR Subtilist; BG13022; yndP.
DR InterPro; IPR000644; CBS_domain.
DR InterPro; IPR005170; CorC_HlyC.
DR - InterPro; IPR002550; DUF21.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01595; DUF21; 1.
DR Pfam; PF03471; CorC_HlyC; 1.
DR SMART; SM00116; CBS; 2.
DR - Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
SQ SEQUENCE 444 AA; 49855 MW; E3A0F57ABE8FB323 CRC64;

Query Match 14.9%; Score 41; DB 1; Length 444;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 62 VDSVXGVFLTNQXD 75
||| : : ||| |
Db 385 VDTGGVFLTNQYD 398

RESULT 11
MYSC DICDI
ID MYSC DICDI STANDARD; PRT; 1181 AA.
AC P42522;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MYOC OR DMIC.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=95348228; PubMed=7622596;
RA Peterson M.D., Novak K.D., Ready M.C., Ruman J.I., Titus M.A.;
RT "Molecular genetic analysis of myoc, a Dictyostelium myosin I.";
RL J. Cell Sci. 108:1093-1103(1995).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED, DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35323; AAC37427.1; -;
DR HSPSP; P08799; IMND.
DR DictyDb; DD01090; myoc.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00063; myosin_head; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin; Actin-binding; ATP-binding; Chemotaxis; SH3 domain;
KW Multigene family.
FT DOMAIN 1 ? MYOSIN HEAD-LIKE.
FT DOMAIN ? 1181 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
FT NP_BIND 109 116 ATP (POTENTIAL).
FT DOMAIN 1122 1181 SH3
SQ SEQUENCE 1181 AA; 132915 MW; 5E1EE47F0CA8803 CRC64;

Query Match 14.9%; Score 41; DB 1; Length 1181;
Best Local Similarity 31.6%; Pred. No. 66;
Matches 6; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 57 WGSVVDVXGVFLTNQXD 75
||| : : ||| |
Db 653 WKGDAISACRAILASQNVYD 671

RESULT 12
RPOB_MYCPN
ID RPOB_MYCPN STANDARD; PRT; 1391 AA.
AC P78013;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
DE beta chain) (RNA polymerase beta subunit).
GN RPOB OR MPN516 OR MP326.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma*
RT *pneumoniae*.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z98682; CAB11357.1; -;
CC EMBL: Z99111; CAB13377.1; -;
CC Subtilist; BG13363; Y1BK.
CC InterPro; IPR002641; Patatin.
CC DR InterPro; IPR001423; UPF0028.
CC PFam; PF01734; Patatin; 1
CC PROSITE; PS01237; UPF0028; 1.
CC Hypothetical protein; Complete proteome.
KW SEQUENCE 260 AA; 28258 MW; 363B6DDF017BE77E CRC64;

Query Match 14.9%; Score 41; DB 1; Length 260;
Best Local Similarity 41.7%; Pred. No. 13;
Matches 10; Conservative 3; Mismatches 5; Indels 6; Gaps 1;

Qy 58 KGSVVD-----SVXGVFLQNXD 75
Db 124 KGSVDAVRASISIPGIFPQRLD 147

RESULT 9
RSMC_BUCAI STANDARD; PRT; 338 AA.
ID RSMC_BUCAI
AC P57413;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52) (rRNA
DE (guanine-N(2)-methyltransferase) (16S rRNA m2G1207
DE methyltransferase).
DE RSCM OR BU328.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
CC NCBI_TaxID:118099;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-Tokyo 1998;
RX MEDLINE-20445173; PubMed-10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT *Buchnera* sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: SPECIFICALLY METHYLATES THE GUANOSINE IN POSITION 1207
CC OF 16S RNA IN THE 30S PARTICLE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-
CC homocysteine + rRNA containing N(2)-methylguanine.
CC -!- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY. RSMC
CC SUBFAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A POTENTIAL FRAMESHIFT

CC WAS CORRECTED IN POSITION 57.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001119; BAB13036.1; ALT_FRAME.
CC InterPro; IPR002052; N6_Mtase.
CC DR InterPro; IPR000051; SAM_bind.
CC PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW rRNA processing; Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 338 AA; 39105 MW; E43516DDDD22FA014 CRC64;

Query Match 14.9%; Score 41; DB 1; Length 338;
Best Local Similarity 36.4%; Pred. No. 18;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 57 WKSVDVSVXGVFLQNXDXXS 78
Db 165 WKNLIKSLPGVFGHKIDSGS 186

RESULT 10
YHDP_BACSU STANDARD; PRT; 444 AA.
ID YHDP_BACSU
AC O07585;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yhdp.
GN YHDP.
OS *Bacillus subtilis*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
CC NCBI_TaxID=1423;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe H., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

CC -I- INDUCTION: Highest expression in darkness.
CC -I- SIMILARITY: BELONGS TO THE AMT1/MEP/NRGA FAMILY OF AMMONIUM
CC TRANSPORTERS (TC 2.49).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF118858; AAG11397.1; -
DR InterPro; IPR001905; Ammonium_transp.
DR Pfam; PF00909; Ammonium_transp; 1.
DR TIGRFAMs; TIGR00836; ant; 1.
KW PROSITE; PS01219; AMMONIUM_TRANS; 1.
KW Transport; Transmembrane; Multigene family.
FT TRANSMEM 15 37 POTENTIAL.
FT TRANSMEM 50 72 POTENTIAL.
FT TRANSMEM 98 117 POTENTIAL.
FT TRANSMEM 124 146 POTENTIAL.
FT TRANSMEM 166 188 POTENTIAL.
FT TRANSMEM 209 227 POTENTIAL.
FT TRANSMEM 255 277 POTENTIAL.
FT TRANSMEM 305 327 POTENTIAL.
FT TRANSMEM 337 356 POTENTIAL.
FT TRANSMEM 377 399 POTENTIAL.
SQ SEQUENCE 460 AA; 49652 MW; 6D9CFC6C2C2395BA CRC64;

Query Match 15.3%; Score 42; DB 1; Length 460;
Best Local Similarity 42.9%; Pred. No. 17;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 57 WKGSVVDVSVXGVFL 70
I: |||:::
DB 5 WEASVTDTSINAIYL 18

RESULT 7
VG69_BPPZA
ID VG69_BPPZA STANDARD; PRT; 108 AA.
AC P08388;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Early protein GP16.9.
GN 16.9.
OS Bacteriophage PZA,
OS Bacteriophage phi-15, and
OS Bacteriophage phi-29.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC phi-29-like viruses.
OX NCBI_TaxID=10757, 10755, 10756;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87031575; PubMed=3095189;
RA Paces V., Vlcek C., Urbanek P., Hostomsky Z.,
RA "Nucleotide sequence of the right early region of Bacillus subtilis
RT phage PZA completes the 19366-bp sequence of PZA genome. Comparison
RT with the homologous sequence of phage phi 29.";
RL Gene 44:115-120(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=phage phi-15;
RX MEDLINE=89232766; PubMed=2497055;
RA Benes V., Arnold L., Sirt J., Paces V.,
RA "Nucleotide sequence of the right early region of Bacillus phage phi
RT 15 and comparison with related phages: reorganization of gene 17
RT during evolution.";
RL Gene 75:341-347(1989).
RN [3]

RP SEQUENCE FROM N.A.
RC SPECIES=Phage phi-29;
RX MEDLINE=86165872; PubMed=3007295;
RA Garvey K.J., Yoshikawa H., Ito J.,
RA "The complete sequence of the Bacillus phage phi 29 right early
RT region.";
RL Gene 40:301-309(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11813; AAA88498.1; -
DR EMBL; M28830; AAA32330.1; -
DR EMBL; M14430; AAA88350.1; -
DR PIR; E29004; WRBP69.
DR PIR; JS0193; WRBP13.
DR PIR; JN0031; JN0031.
KW Early protein.
SQ SEQUENCE 108 AA; 12648 MW; 4FA9C051B60CD8F9 CRC64;

Query Match 14.9%; Score 41; DB 1; Length 108;
Best Local Similarity 35.0%; Pred. No. 5.3;
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 52 RFXXXKGSVVDVSVXGVFLT 71
I: |||:::
DB 59 RRFASWSQGLGGSNSVFWT 78

RESULT 8
YLBK_BACSU
ID YLBK_BACSU STANDARD; PRT; 260 AA.
AC O34731;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ylbk.
GN YLBK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Bouilliet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haleth J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

Query Match 15.6%; Score 43; DB 1; Length 474;
Best Local Similarity 58.3%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 57 WKGSVVDSVXGV 68
I:||||:|I:|
DB 89 WEGSVLDSEEGI 100

RESULT 4

CISY_ASPNG STANDARD; PRT; 475 AA.
AC P51044;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).
GN CIT-1.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WU-2223L;
RA Oshida Y., Miyake K., Kanayama S., Kirimura K., Usami S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H(2)O +
oxaloacetate.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D63376; BAA09691.1; -;
DR HSSP; P23007; 2CSC.
DR InterPro; IPR002020; Citrate_synt.
DR Pfam; PF00285; citrate_synt.1.
DR PRINTS; PR00143; CITRNTNTHASE
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
KW Lyase; Tricarboxylic acid cycle; Mitochondrion; Transist peptide.
FT TRANSIT 1 ? MITOCHONDRION (BY SIMILARITY).
FT CHAIN ? 475 CITRATE SYNTHASE.
FT ACT_SITE 310 310 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
SQ SEQUENCE 475 AA; 52153 MW; F932525B3F31FCB3F CRC64;

Query Match 15.6%; Score 43; DB 1; Length 475;
Best Local Similarity 58.3%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 57 WKGSVVDSVXGV 68
I:||||:|I:|
DB 89 WEGSVLDSEEGI 100

RESULT 5

SPIH_HUMAN STANDARD; PRT; 232 AA.
AC Q99865; O75650; Q9UJ0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Spindlin homolog (Protein DXF34).
GN DXF34.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97419273; PubMed=9271673;
RA Laval S.H., Reed V., Blair H.J., Boyd Y.;
RT "The structure of DXF34, a human X-linked sequence family with
RT homology to a transcribed mouse Y-linked repeat.";
RL Mamm. Genome 8:689-691(1997).
RN [2]

RP SEQUENCE FROM N.A.
RA Heath P.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SPIN / STSY FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; Y09858; CAA70988.1; -;
DR EMBL; AL022157; CAA18148.1; -;
DR EMBL; AL022157; CAA18149.1; -;
DR InterPro; IPR003671; Spin_Ssty.
DR Pfam; PF02513; Spin-Ssty; 3.
KW Developmental protein; Phosphorylation; Cell cycle.
FT MOD_RES 196 196 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 224 224 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 20 20 C -> R (IN REF. 2).
FT CONFLICT 208 208 T -> A (IN REF. 2; CAA18149).
SQ SEQUENCE 232 AA; 26536 MW; 093E7E9F5340BF71 CRC64;

Query Match 15.3%; Score 42; DB 1; Length 232;
Best Local Similarity 41.2%; Pred. No. 8.1;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 49 QGXRFXKXKGVSDV 65
I:||||:|I:|
DB 34 EGDEPITQWKGTVLDQV 50

RESULT 6

AT13_LYCES STANDARD; PRT; 460 AA.
AC Q9FVNO;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ammonium transporter 1, member 3 (Leamt1;3).
GN AMT1.3.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=20208060; PubMed=10743657;
RA von Wiren N., Lauter F.-R., Ninnemann O., Gillissen B., Walch-Liu P.,
RA Engels C., Jost W., Frommer W.B.;
RT "Differential regulation of three functional ammonium transporter
RT genes by nitrogen in root hairs and by light in leaves of tomato.";
RL Plant J. 21:167-175(2000).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (probable).
CC -!- TISSUE SPECIFICITY: Leaves.

```
RL Nature 390:249-256(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84432; BAA12531.1; -.
DR EMBL; Z99116; CAB14406.1; -.
DR Subtilist; BG11695; yqba.
DR InterPro; IPR000644; CBS_domain.
DR InterPro; IPR005170; CorC_HlyC.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01595; DUF21; 1.
DR Pfam; PF03471; CorC_HlyC; 1.
KW - Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 16 36 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
SQ -SEQUENCE 442 AA; 49956 MW; E2EA41573549DA83 CRC64;

Query Match 17.1%; Score 47; DB 1; Length 442;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 62 VDSVXGFLTQNXSSXA 81
II:: I I I I I I I I I I
DB 396 VDTLGGWFLTQNDAPESA 415

RESULT 2
CISY_CANTR
ID CISY_CANTR STANDARD; PRT; 467 AA.
AC P79024;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).
GN CIT.
OS Candida tropicalis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5482;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PK233.
RA Ueda M., Sanuki S., Kawachi H., Shimizu K., Atomi H., Tanaka A.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H(2)O +
CC oxaloacetate.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CC CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001565; BAA19410.1; -.
DR
```

```
DR HSSP; P23007; 2CSC.
DR InterPro; IPR002020; Citrate_synt.
DR Pfam; PF00285; citrate_synt; 1.
DR PRINTS; PR00143; CITRATNTNTHASE.
DR PROSITE; PS00480; CITRATE-SYNTHASE; 1.
KW Lyase; Tricarboxylic acid cycle; MITOCHONDRION (POTENTIAL).
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN 1 467 CITRATE SYNTHASE.
FT ACT_SITE 301 301 BY SIMILARITY.
FT ACT_SITE 347 347 BY SIMILARITY.
FT ACT_SITE 402 402 BY SIMILARITY.
SQ SEQUENCE 467 AA; 52004 MW; 4B194132C4198CA2 CRC64;

Query Match 15.6%; Score 43; DB 1; Length 467;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 57 WKGSVVDVXGV 68
II:: I I I I I I I I I I
DB 79 WEGSVLDPIEGI 90

RESULT 3
CISY_EMENI
ID CISY_EMENI STANDARD; PRT; 474 AA.
AC O00098;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).
GN CIT.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=97306446; PubMed=9163747;
RA Park B.W., Han K.H., Lee C.Y., Lee C.H., Maeng P.J.;
RA "Cloning and characterization of the cita gene encoding the
RA mitochondrial citrate synthase of Aspergillus nidulans."
RA Mitochondrial citrate synthase of Aspergillus nidulans.
RL Mol. Cells 7:290-295(1997).
CC -!- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H(2)O +
CC oxaloacetate.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CC CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U89675; AAC49728.1; -.
DR HSSP; P23007; 2CSC.
DR InterPro; IPR002020; Citrate_synt.
DR Pfam; PF00285; citrate_synt; 1.
DR PRINTS; PR00143; CITRATNTNTHASE.
DR PROSITE; PS00480; CITRATE-SYNTHASE; 1.
KW Lyase; Tricarboxylic acid cycle; MITOCHONDRION (POTENTIAL).
FT TRANSIT 1 35 MITOCHONDRION (POTENTIAL).
FT CHAIN 36 474 CITRATE SYNTHASE.
FT ACT_SITE 310 310 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
SQ SEQUENCE 474 AA; 52223 MW; E2E86892ACB5398B CRC64;
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:14 ; Search time 3.23529 Seconds
(without alignments)
1153.796 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KXXXXXXXXXXXXXD.....TQNXDXXSXAXXXXXFP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	17.1	442	1 YQHB_BACSU	P54505 bacillus su
2	43	15.6	467	1 C1SY_CANTR	P79024 candida tro
3	43	15.6	474	1 C1SY_EMENI	O00098 emericella
4	43	15.6	475	1 C1SY_ASPNG	P51044 aspergillus
5	42	15.3	232	1 SPIH_HUMAN	Q99865 homo sapien
6	42	15.3	460	1 ATR13_LYCES	Q9fiv0 lycopersico
7	41	14.9	108	1 V669_BPPZA	P08388 bacterioph
8	41	14.9	260	1 YLBK_BACSU	O34731 bacillus su
9	41	14.9	338	1 RSMC_BUCAI	P57413 buchnera ap
10	41	14.9	444	1 YHDP_BACSU	O07585 bacillus su
11	41	14.9	1181	1 MYSC_DICDI	P42522 dictyosteli
12	41	14.9	1391	1 RPOB_MYCPN	P78013 mycoplasma
13	40	14.5	191	1 CLD7_RAT	Q92111 rattus norv
14	40	14.5	211	1 CLD7_MOUSE	Q92261 mus musculu
15	40	14.5	318	1 DEDD_HUMAN	O75618 homo sapien
16	40	14.5	318	1 DEDD_MOUSE	Q92213 mus musculu
17	40	14.5	318	1 DEDD_RAT	Q922k0 rattus norv
18	40	14.5	469	1 C1SY_NEUCR	P34085 neurospora
19	40	14.5	796	1 YFS7_CABEL	Q10003 caenorhabdi
20	40	14.5	960	1 CHPL_SCHPO	Q10103 schizosacch
21	40	14.5	1503	1 MRP6_HUMAN	Q95255 homo sapien
22	40	14.5	1526	1 YK46_ANASP	Q9yrl1 anabaena sp
23	40	14.5	3911	1 AKA9_HUMAN	Q99996 h a-kinase
24	39.5	14.4	164	1 ATPF_STRPN	Q59952 streptococc
25	39	14.2	208	1 CT30_ARATH	Q91q17 arabidopsis
26	39	14.2	237	1 SPIN_HUMAN	Q9y657 homo sapien
27	39	14.2	240	1 SPIN_MOUSE	P61142 mus musculu
28	39	14.2	268	1 COX3_SCHCO	P14058 schizophyil
29	39	14.2	270	1 YF53_AQUAE	O67438 aquifex aeo
30	39	14.2	313	1 EBAG_STRPL	P04067 streptomyc
31	39	14.2	314	1 YG20_YEAST	P53257 saccharomyc
32	39	14.2	373	1 DP3B_MYCPU	Q98rk6 mycoplasma
33	39	14.2	389	1 UTL1_HUMAN	Q13336 homo sapien

RESULT 1

ID	YQHB_BACSU	STANDARD:	PRT:	442 AA.
DT	P54505;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yqnb.			
GN	YQHB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168 / JH642;			
RX	MEDLINE=97124195; PubMed=8969508;			
RA	Mizuno M., Masuda S., Takemaru K.I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;			
RT	"Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes."			
RL	Microbiology 142:3103-3111(1996).			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tatesato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus subtilis."			

ALIGNMENTS

THIS PAGE BLANK (USPTO)

Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
DNA demethylation; transgenic plant; transcription modulation;

Search completed: November 8, 2002, 19:26:00
Job time : 48.3268 secs


```
XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX PF WPI; 2002-055307/07.
XX DR N-PSDB; AAS96709.
XX PR
XX PA New polynucleotide that control plant development comprising a sequence
XX PT having a specific homology to DEMETER domains A,B or C
XX PS Disclosure; Page 94; 109pp; English.
XX
XX CC The invention relates to an isolated polynucleotide sequence or their
XX CC complement encoding a polypeptide having a sequence at least 40%
XX CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
XX CC or C or their combinations. Also included are an expression
XX CC cassette comprising the polynucleotide or comprising a heterologous
XX CC polynucleotide under the control of a promoter at least 70% identical to
XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
XX CC region of DMT, a host cell comprising an exogenous polynucleotide
XX CC encoding a DMT-like protein and a transgenic plant comprising a
XX CC polynucleotide encoding a DMT-like protein. The expression cassette is
XX CC useful for modulating transcription. The method comprises introducing
XX CC the cassette into a host cell preferably Agrobacterium by sexual
XX CC cross, and selecting a host cell with modulated transcription, where
XX CC the protein is capable of exhibiting at least one of the following
XX CC biological activities, which include enhanced expression of the
XX CC protein in a plant results in a delay in flowering time, introduction of
XX CC chromosomal DNA in the cell, reduction of expression of the protein in a
XX CC plant results in enhanced endosperm development and expressing of the
XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.
XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates
XX CC transcription of target genes by demethylation. The present
XX CC sequence represents a DMT-like protein.
XX SQ Sequence 164 AA;
Query Match 14.7%; Score 87; DB 23; Length 164;
Best Local Similarity 22.7%; Pred. No. 3.5e-10;
Matches 34; Conservative 0; Mismatches 112; Indels 4; Gaps 2;
Qy 133 YLLXIWPXXXXXXXXXXXXXXXXXXXXXADHXXXXXXXXXXXXXXXXXXXXXRGTXLX 192
Db 16 YLLSIWTP-GETAOSIDAPKTFCT---DSGETGRLCGSGTSCFSCNNIREMQAKVGRGTLI 71
Qy 193 XXXXXXXXXXXXXXXXXXXXXXXXADHXXXXXXXXXXPIXXXXXXXXXXXXXGXXXXXXX 252
Db 72 PCRTAMRGSPFLNGTYFOVNEVFADHCSQNPIDVPRSWIWDLPRTVYFGTSVPTIFRG 131
Qy 253 LXXXXXXXFXGXXCXRFXXXXXXXXXXL 282
Db 132 LTTEEIQCWFVRGFCVGRGDTVRAPRAL 161
RESULT 13
AAU72764
ID AAU72764 standard; Protein; 179 AA.
XX
XX AC AAU72764;
XX
XX DT 26-FEB-2002 (first entry)
XX DE Corn DMT protein related sequence #2.
XX
XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
XX KW DNA demethylation; transgenic plant; transcription modulation;
XX KW flowering time; endosperm development; MEDEA.
XX OS Zea mays.
XX
XX PN WO200180626-A1.
XX
```

```
PD 01-NOV-2001.
XX
XX PF 23-APR-2001; 2001WO-US13059.
XX
XX PR 21-APR-2000; 2000US-0553690.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX
XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
XX DR WPI; 2002-055307/07.
XX DR N-PSDB; AAS96714.
XX
XX PT New polynucleotide that control plant development comprising a sequence
XX PT having a specific homology to DEMETER domains A,B or C
XX PS Disclosure; Page 97; 109pp; English.
XX
XX CC The invention relates to an isolated polynucleotide sequence or their
XX CC complement encoding a polypeptide having a sequence at least 40%
XX CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
XX CC or C or their combinations. Also included are an expression
XX CC cassette comprising the polynucleotide or comprising a heterologous
XX CC polynucleotide under the control of a promoter at least 70% identical to
XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
XX CC region of DMT, a host cell comprising an exogenous polynucleotide
XX CC encoding a DMT-like protein and a transgenic plant comprising a
XX CC polynucleotide encoding a DMT-like protein. The expression cassette is
XX CC useful for modulating transcription. The method comprises introducing
XX CC the cassette into a host cell preferably Agrobacterium by sexual
XX CC cross, and selecting a host cell with modulated transcription, where
XX CC the protein is capable of exhibiting at least one of the following
XX CC biological activities, which include enhanced expression of the
XX CC protein in a plant results in a delay in flowering time, introduction of
XX CC the protein into a cell results in modulation of methylation of
XX CC chromosomal DNA in the cell, reduction of expression of the protein in a
XX CC plant results in enhanced endosperm development and expressing of the
XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.
XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates
XX CC transcription of target genes by demethylation. The present
XX CC sequence represents a DMT-like protein.
XX SQ Sequence 179 AA;
Query Match 14.4%; Score 85; DB 23; Length 179;
Best Local Similarity 23.3%; Pred. No. 1e-09;
Matches 37; Conservative 0; Mismatches 118; Indels 4; Gaps 2;
Qy 129 DXXYLLXIWPXXXXXXXXXXXXXXXXXXXXXADHXXXXXXXXXXXXXCCXXXXXRG 188
Db 18 DPTPYLLAIWTP--DGKEITKTPKPC--CDPQMGGDLCNNEMCHNCTAEKENQSRVVRG 73
Qy 189 TXLXXXXXXXXXXXXXXXXXXXXXADHXXXXXXXXXXPIXXXXXXXXXXXXXRG 248
Db 74 TILVPCRTAMRGSPFLNGTYFOVNEVFADHRSSSHNPINHVIERMLNQLQRMVFFGTSVPT 133
Qy 249 IXXLXXXXXXXXXXXXXGXXCXRFXXXXXXXXXXXXXRXLXLH 287
Db 134 IFKGLRTEEQCFWRGFCVGRGFDMTAPRPLCPHLH 172
RESULT 14
AAU72754
ID AAU72754 standard; Protein; 209 AA.
XX
XX AC AAU72754;
XX
XX DT 26-FEB-2002 (first entry)
XX DE Soybean DMT3 protein.
XX
XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
```

```
PT New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
PS Disclosure; Page 91; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMTET, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents a DMT-like protein.
XX
SQ Sequence 102 AA;

Query Match 15.0%; Score 88.5; DB 23; Length 102;
Best Local Similarity 28.3%; Pred. No. 1e-10;
Matches 28; Conservative 0; Mismatches 52; Indels 19; Gaps 3;

QY 104 KXXXXLRTHEHXXVXXLPDXHXXLXXDX-----XXYLLXIXWXPXXXXX-----XXXXXX 152
| | | | | | | | | | | | | | | | | | | | |
DB 11 KNVSLRTEHCYVYELPDHPHLLQGMWDTREPDGPKGLAIWTPGETANSIQPPESKCSQ 70
| | | | | | | | | | | | | | | | | | | | |
QY 153 XXCXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 191
| | | | | | | | | | | | | | | | | | | | |
DB 71 EEC-----GQLCNEECFSCNSFRANSQIVRGTL 101
| | | | | | | | | | | | | | | | | | | | |

RESULT 11
AAU72760
ID AAU72760 standard; Protein; 137 AA.
XX
AC AAU72760;
XX
DT 26-FEB-2002 (first entry)
XX
DE Cotton DMT protein related sequence.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Gossypium hirsutum.
XX
PN W0200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX
```

```
DR WPI: 2002-055307/07.
DR N-PSDB; AAS96710.
XX
PT New polynucleotide that control plant development comprising a sequence
PT having a specific homology to DEMETER domains A,B or C
PS Disclosure; Page 94; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMTET, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
CC polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
CC cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
CC biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
CC the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
CC plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
CC transcription of target genes by demethylation. The present
CC sequence represents a DMT-like protein.
XX
SQ Sequence 137 AA;

Query Match 14.7%; Score 87; DB 23; Length 137;
Best Local Similarity 29.8%; Pred. No. 2.9e-10;
Matches 28; Conservative 0; Mismatches 56; Indels 10; Gaps 3;

QY 104 KXXXXLRTHEHXXVXXLPDXHXXLXX-----XDXXXYLLXIXWXPXXXXX-----XX 157
| | | | | | | | | | | | | | | | | | | | |
DB 32 KNVSLRTEHCYVYELPDHPHLLQGMWDTREPDGPKGLAIWTPGETANSIQPPESQSC 157
| | | | | | | | | | | | | | | | | | | | |
QY 158 XXXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 191
| | | | | | | | | | | | | | | | | | | | |
DB 89 -GSQEPGRCLNEKTCFACNSVREANTETVRGTIL 121
| | | | | | | | | | | | | | | | | | | | |

RESULT 12
AAU72759
ID AAU72759 standard; Protein; 164 AA.
XX
AC AAU72759;
XX
DT 26-FEB-2002 (first entry)
XX
DE Corn DMT protein related sequence #1.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Zea mays.
XX
PN W0200180626-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC ) UNIV CALIFORNIA.
```


CC	cassette comprising the polynucleotide or comprising a heterologous
CC	polynucleotide under the control of a promoter at least 70% identical to
CC	DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC	region of DMT, a host cell comprising an exogenous polynucleotide
CC	encoding a DMT-like protein and a transgenic plant comprising a
CC	polynucleotide encoding a DMT-like protein. The expression cassette is
CC	useful for modulating transcription. The method comprises introducing
CC	the cassette into a host cell preferably Agrobacterium by sexual
CC	cross, and selecting a host cell with modulated transcription, where
CC	the protein is capable of exhibiting at least one of the following
CC	biological activities, which include enhanced expression of the
CC	protein in a plant results in a delay in flowering time, introduction of
CC	the protein into a cell results in modulation of methylation of
CC	chromosomal DNA in the cell, reduction of expression of the protein in a
CC	plant results in enhanced endosperm development and expressing of the
CC	protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC	The polynucleotide is useful for detecting a nucleic acid in a sample.
CC	DEMETER is related to 5-methylcytosine glycosylases and regulates
CC	transcription of target genes by demethylation. The present
CC	sequence represents a DMT-like protein.
XX	
SQ	Sequence 210 AA;
	Query Match 22.4%; Score 132; DB 23; Length 210;
	Best Local Similarity 24.9%; Pred. No. 1.le-19;
	Matches 47; Conservative 0; Mismatches 124; Indels 18; Gaps 3;
QY	104 KXXXXLRTHEHVVXLLPDHXHLLX-----XDXXXXVLLXIWXPKXXXXX----XXXXXXXX 153
Db	28 KNVSLRTEHSVEYPDSHPLEGGWKEKREPPDPKGKLLAIWTGGETANSTQPPDRRCSAQ 87
QY	154 XCXXXXXXXKXCXKX 213
Db	88 DC-----GQLCNEEECFSCNSFREANSOIVRGTLILPCRTAMRGSFPLNGTYFQVNE 139
QY	214 XFADHXXXXXPXXXXXXXKX 273
Db	140 VFADHESLNPIVSPSRLIWNLDRTVHFGTSVTSIFKGLATPETQOCFWRGFVCVRSE 199
QY	274 XXXXXPPXL 282
Db	200 RSTRAPRPL 208
	RESULT 8
	ABB92260
ID	ABB92260 standard; Protein; 1309 AA.
XX	ABB92260;
XX	
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Herbicidally active polypeptide SEQ ID NO 1471.
XX	
KW	Herbicidal; plant; agriculture; herbicide.
OS	Arabidopsis thaliana.
XX	
PN	WO200210210-A2.
XX	
PD	07-FEB-2002.
XX	
PF	28-AUG-2001; 2001WO-EPO9892.
XX	
PR	28-AUG-2001; 2001WO-EPO9892.
XX	
PA	(FARB) BAYER AG.
XX	
PI	Tietjen K, Weidler M;
XX	
DR	WPI; 2002-269010/31.
XX	
PT	Identifying plant target proteins for herbicidally active compounds.

KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Oryza sativa.
XX WO200180626-A1.
XX 01-NOV-2001.
XX 23-APR-2001; 2001WO-US13059.
XX 21-APR-2000; 2000US-0553690.
XX (REGC) UNIV CALIFORNIA.
XX Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX WPI; 2002-055307/07.
XX N-PSDB; AAS96695.
XX New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
XX
XX Disclosure; Page 79-80; 109pp; English.
XX The invention relates to an isolated polynucleotide sequence or their
XX complement encoding a polypeptide having a sequence at least 40%
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
XX or C or their combinations. Also included are an expression
XX cassette comprising the polynucleotide or comprising a heterologous
XX polynucleotide under the control of a promoter at least 70% identical to
XX DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
XX region of DMT, a host cell comprising an exogenous polynucleotide
XX encoding a DMT-like protein and a transgenic plant comprising a
XX polynucleotide encoding a DMT-like protein. The expression cassette is
XX useful for modulating transcription. The method comprises introducing
XX the cassette into a host cell preferably Agrobacterium by sexual
XX cross, and selecting a host cell with modulated transcription, where
XX the protein is capable of exhibiting at least one of the following
XX biological activities, which include enhanced expression of the
XX protein in a plant results in a delay in flowering time, introduction of
XX the protein into a cell results in modulation of methylation of
XX chromosomal DNA in the cell, reduction of expression of the protein in a
XX plant results in enhanced endosperm development and expression of the
XX protein in an Arabidopsis leaf results in expression of the MEDEA gene.
XX The polynucleotide is useful for detecting a nucleic acid in a sample.
XX DEMETER is related to 5-methylcytosine glycosylases and regulates
XX transcription of target genes by demethylation. The present
XX sequence represents a DMT-like protein.
XX
SQ Sequence 1952 AA;
Query Match 24.6%; Score 145; DB 23; Length 1952;
Best Local Similarity 26.2%; Pred. No. 1.8e-21;
Matches 51; Conservative 0; Mismatches 134; Indels 10; Gaps 3;
QY 104 KXXXLRTHEHXXVXXLPDXHXXLXX-----XDXXXYLXIXWXPXXXXXXXCX 157
DB 1745 KXVSLRTEHQVYELPSHPLLEGFNQREPDPCPYLLSIWTP-CETAQSDAPKSV-- 1801
QY 158 XXXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 217
DB 1802 -NSQNGELCASNTFCSCNSTREAQAQVGTGLLIPCTAMRGSPPLNGTVFQVNEVAD 1860
QY 218 HXXXXXPXXXXXXXLXRXXXXXXXXLXIXXXXXXXXLXIXXXXXXXXLXIXXXXX 277
DB 1861 HDSSRNPDVPRSWLWNLPRRTVFGTSGIPFGLTTEIHOCHFWRGFCVGRGFDTSR 1920
QY 278 XPXXLXXLHXXXXX 292
DB 1921 APRPLYARLHPASK 1935

RESULT 4
AAU72756
ID AAU72756 standard; Protein; 217 AA.
XX
AC AAU72756;
XX
DT 26-FEB-2002 (first entry)
XX
DE Alfalfa DMT protein related sequence.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Medicago sativa.
XX WO200180626-A1.
XX
XX 01-NOV-2001.
XX 23-APR-2001; 2001WO-US13059.
XX 21-APR-2000; 2000US-0553690.
XX (REGC) UNIV CALIFORNIA.
XX Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
XX WPI; 2002-055307/07.
XX N-PSDB; AAS96706.
XX New polynucleotide that control plant development comprising a sequence
XX having a specific homology to DEMETER domains A,B or C
XX
XX Disclosure; Page 92; 109pp; English.
XX The invention relates to an isolated polynucleotide sequence or their
XX complement encoding a polypeptide having a sequence at least 40%
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
XX or C or their combinations. Also included are an expression
XX cassette comprising the polynucleotide or comprising a heterologous
XX polynucleotide under the control of a promoter at least 70% identical to
XX DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
XX region of DMT, a host cell comprising an exogenous polynucleotide
XX encoding a DMT-like protein and a transgenic plant comprising a
XX polynucleotide encoding a DMT-like protein. The expression cassette is
XX useful for modulating transcription. The method comprises introducing
XX the cassette into a host cell preferably Agrobacterium by sexual
XX cross, and selecting a host cell with modulated transcription, where
XX the protein is capable of exhibiting at least one of the following
XX biological activities, which include enhanced expression of the
XX protein in a plant results in a delay in flowering time, introduction of
XX the protein into a cell results in modulation of methylation of
XX chromosomal DNA in the cell, reduction of expression of the protein in a
XX plant results in enhanced endosperm development and expression of the
XX protein in an Arabidopsis leaf results in expression of the MEDEA gene.
XX The polynucleotide is useful for detecting a nucleic acid in a sample.
XX DEMETER is related to 5-methylcytosine glycosylases and regulates
XX transcription of target genes by demethylation. The present
XX sequence represents a DMT-like protein.
XX
SQ Sequence 217 AA;
Query Match 24.2%; Score 143; DB 23; Length 217;
Best Local Similarity 25.3%; Pred. No. 4.9e-22;
Matches 50; Conservative 0; Mismatches 130; Indels 18; Gaps 3;
QY 104 KXXXLRTHEHXXVXXLPDXHXXLXX-----XDXXXYLXIXWXPXXXXXXX 153
DB 28 KXVSLRTEHQVYELPSHPLLEGFNQREPDPCPYLLSIWTPGFTANSIQPPDRCSAQ 87
QY 154 XXXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 213

XX PS Claim 3; Page 54-55; 109pp; English.

XX CC The invention relates to an isolated polynucleotide sequence or their

XX CC complement encoding a polypeptide having a sequence at least 40%

XX CC identical to DMT (DEMTET, previously known as ATROPOS (ATR)) Domain A, B

XX CC or C or their combinations. Also included are an expression

XX CC cassette comprising the polynucleotide or comprising a heterologous

XX CC polynucleotide under the control of a promoter at least 70% identical to

XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

XX CC region of DMT, a host cell comprising an exogenous polynucleotide

XX CC encoding a DMT-like protein and a transgenic plant comprising a

XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

XX CC useful for modulating transcription. The method comprises introducing

XX CC the cassette into a host cell preferably Agrobacterium by sexual

XX CC cross, and selecting a host cell with modulated transcription, where

XX CC the protein is capable of exhibiting at least one of the following

XX CC biological activities, which include enhanced expression of the

XX CC protein in a plant results in a delay in flowering time, introduction of

XX CC the protein into a cell results in modulation of methylation of

XX CC chromosomal DNA in the cell, reduction of expression of the protein in a

XX CC plant results in enhanced endosperm development and expressing of the

XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.

XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.

XX CC DEMTER is related to 5-methylcytosine glycosylases and regulates

XX CC transcription of target genes by demethylation. The present

XX CC sequence represents a DMT-like protein.

XX SQ Sequence 1729 AA;

Query Match 24.7%; Score 146; DB 23; Length 1729;

Best Local Similarity 26.2%; Pred. No. 9.6e-22;

Matches 51; Conservative 0; Mismatches 134; Indels 10; Gaps 3;

Qy 104 KXXXLRTHEHVXXLPDXHXXLXXDX-----XXYLXIXWXPXXXXXXXCCXX 157

Db 1533 KNISRLRTHQVYELPDEHRLDGMKREPDPDPSYLLAIWTP-GETANSQPEQKC-- 1589

Qy 158 XXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 217

Db 1590 -GGKASGKMFDETCSECNLSREANSQTVRGTLIPCRMTAMRGSPFLNGTYFQVNEFPAD 1648

Qy 218 HXXXXXXPTXXXXXXLXXRXXXXGXXXXXXLXXLXXLXXLXXLXXLXXLXXLXX 277

Db 1649 HESSLKPIDVPRDMWDLPRRTVYFGTSVTSIFRGLSTEQIOFCFWKGFVGVCFEQKTR 1708

Qy 278 XPXXLXXLHXXSK 292

Db 1709 APRPLMARLHFPASK 1723

RESULT 2

AAU72737

ID AAU72737 standard; Protein; 1413 AA.

XX AC

XX DT 26-FEB-2002 (first entry)

XX XX

XX DE Arabidopsis DMT2 (IDMT2) protein.

XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;

XX KW DNA demethylation; transgenic plant; transcription modulation;

XX KW flowering time; endosperm development; MEDEA.

XX OS Arabidopsis Thaliana.

XX PN WO200180626-A1.

XX XX

XX PD 01-NOV-2001.

XX XX

XX PF 23-APR-2001; 2001WO-US13059.

XX KW

PR 21-APR-2000; 2000US-0553690.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;

XX XX

XX WPI; 2002-055307/07.

DR N-PSDB; AAS96692.

XX XX

XX New polynucleotide that control plant development comprising a sequence

XX PT having a specific homology to DEMETER domains A,B or C

XX PS Disclosure; Page 67; 109pp; English.

XX CC The invention relates to an isolated polynucleotide sequence or their

XX CC complement encoding a polypeptide having a sequence at least 40%

XX CC identical to DMT (DEMTET, previously known as ATROPOS (ATR)) Domain A, B

XX CC or C or their combinations. Also included are an expression

XX CC cassette comprising the polynucleotide or comprising a heterologous

XX CC polynucleotide under the control of a promoter at least 70% identical to

XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

XX CC region of DMT, a host cell comprising an exogenous polynucleotide

XX CC encoding a DMT-like protein and a transgenic plant comprising a

XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

XX CC useful for modulating transcription. The method comprises introducing

XX CC the cassette into a host cell preferably Agrobacterium by sexual

XX CC cross, and selecting a host cell with modulated transcription, where

XX CC the protein is capable of exhibiting at least one of the following

XX CC biological activities, which include enhanced expression of the

XX CC protein in a plant results in a delay in flowering time, introduction of

XX CC the protein into a cell results in modulation of methylation of

XX CC chromosomal DNA in the cell, reduction of expression of the protein in a

XX CC plant results in enhanced endosperm development and expressing of the

XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.

XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.

XX CC DEMTER is related to 5-methylcytosine glycosylases and regulates

XX CC transcription of target genes by demethylation. The present

XX CC sequence represents a DMT-like protein.

XX SQ Sequence 1413 AA;

Query Match 24.6%; Score 145; DB 23; Length 1413;

Best Local Similarity 26.2%; Pred. No. 1.3e-21;

Matches 51; Conservative 0; Mismatches 134; Indels 10; Gaps 3;

Qy 104 KXXXLRTHEHVXXLPDXHXXLXXDX-----XXYLXIXLXWXPXXXXXXXCCXX 157

Db 1214 KNISQLRTHRVYELPDEHPLLAOLEKREPDDPCSYLLAIWTP-GETADSIQSVSTC-- 1270

Qy 158 XXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 217

Db 1271 -IFOANGMLCDEETCFSCNSIKETRSQIVRGTLIPCRMTAMRGSPFLNGTYFQVNEFPAD 1329

Qy 218 HXXXXXXPTXXXXXXLXXRXXXXGXXXXXXLXXLXXLXXLXXLXXLXXLXXLXX 277

Db 1330 HASLNPINVPRELIMELPRRTVYFGTSVPTIFKGLSTEKIQACFWKGFVGVCFDRKTR 1389

Qy 278 XPXXLXXLHXXSK 292

Db 1390 GPKPLIARLHFPASK 1404

RESULT 3

AAU72743

ID AAU72743 standard; Protein; 1952 AA.

XX AC

XX AC AAU72743;

XX XX

XX DT 26-FEB-2002 (first entry)

XX XX

XX DE Rice DMT1 protein.

XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:08 ; Search time 45.3268 Seconds
(without alignments)
858.415 Million cell updates/sec

Title: US-09-840-743-73
Perfect score: 590
Sequence: 1 PXEXPPXXXXX.....XXXXXXLXXLHXXXXX 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Genes2_101002.*
1: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	24.7	1729	23 AAU72736	Arabidopsis Demete
2	145	24.6	1413	23 AAU72737	Arabidopsis DMT2 (
3	145	24.6	1952	23 AAU72743	Rice DMT1 protein.
4	143	24.2	217	23 AAU72756	Alfalfa DMT protei
5	137.5	23.3	1332	23 AAU72739	Arabidopsis DMT3 (
6	134.5	22.8	1114	23 AAU72741	Arabidopsis DMT4 (
7	132	22.4	210	23 AAU72757	Tomato DMT protein
8	108.5	18.4	1309	23 ABB92260	Herbicidally activ
9	100	16.9	177	23 AAU72745	Corn DMT1 protein.
10	88.5	15.0	102	23 AAU72755	Soybean DMT4 prote

11	87	14.7	137	23 AAU72760	Cotton DMT protein
12	87	14.7	164	23 AAU72759	Corn DMT protein r
13	85	14.4	179	23 AAU72764	Corn DMT protein r
14	79	13.4	209	23 AAU72754	Soybean DMT3 prote
15	77	13.1	108	23 AAU72750	Wheat DMT3 protein
16	77	13.1	140	23 AAU72766	Corn DMT protein r
17	76	12.9	138	23 AAU72767	Corn DMT protein r
18	74	12.5	150	23 AAU72765	Corn DMT protein r
19	72	12.2	89	23 AAU72768	Corn DMT protein r
20	66	11.2	93	23 AAU72769	Corn DMT protein r
21	59	10.0	183	23 AAU72748	Wheat DMT1 protein
22	54	9.2	71	23 AAU72770	Corn DMT protein r
23	41	6.9	1548	22 ABB62738	Drosophila melanog
24	41	6.9	1548	22 AAB72443	UGR. Drosophila
25	41	6.9	1548	23 AAU72583	Drosophila cell cy
26	37	6.3	229	21 AAU90301	Human peptidase, H
27	37	6.3	229	22 AAU12260	Human PRO4389 poly
28	37	6.3	233	22 AAU92597	Human digestive sy
29	37	6.3	233	22 AAM38606	Human colorectal c
30	37	6.3	3503	22 ABB66499	Drosophila melanog
31	36	6.1	85	22 AAM88162	Human immune/haema
32	36	6.1	144	21 AAG35357	Zea mays protein f
33	36	6.1	204	22 AAU50869	Propionibacterium
34	36	6.1	662	21 AAU95685	Cosmid cHRIM5 enco
35	35	5.9	75	22 AAU48812	Propionibacterium
36	35	5.9	81	22 AAU53845	Propionibacterium
37	35	5.9	606	20 AAU43219	E. thusiopathiae s
38	34.5	5.8	1289	22 ABB70840	Drosophila melanog
39	34	5.8	49	21 AAG27372	Arabidopsis thalia
40	34	5.8	83	22 AAU22769	Human prostate can
41	34	5.8	83	22 AAU94749	Human reproductive
42	34	5.8	109	21 AAG06711	Arabidopsis thalia
43	34	5.8	112	21 AAG06710	Arabidopsis thalia
44	34	5.8	152	23 AAU10493	Human HOMO tyrosin
45	34	5.8	154	22 AAB60106	Human transport pr

ALIGNMENTS

RESULT 1
AAU72736
ID AAU72736 standard; Protein: 1729 AA.
XX AAU72736;
AC AC
XX XX
DT 26-FEB-2002 (first entry)
XX XX
DE Arabidopsis Demeter protein.
XX XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX XX
OS Arabidopsis Thaliana.
XX OS
PN WO200180626-A1.
XX XX
PD 01-NOV-2001.
XX XX
PF 23-APR-2001; 2001WO-US13059.
XX XX
PR 21-APR-2000; 2000US-0553690.
XX XX
XX (REGC) UNIV CALIFORNIA.
XX PA
XX Arabidopsis DMT2 (

PI Fischer RL, Choi Y, Hannon M, Okumuro JK, Tatarinova TV;
XX WPI; 2002-055307/07.
DR N-PSDB; AAS96686, AAS96689.
XX XX
PT New polynucleotide that control plant development comprising a sequence having a specific homology to DEMETER domains A, B or C -

THIS PAGE BLANK (USPTO)

; REFERENCE/DOCKET NUMBER: DX0335K1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-110-683-2

Query Match 5.43; Score 32; DB 1; Length 578;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 109 LRTEHXXVXXLPDXH 122
I:|I| |I|I|
Db 328 LQTEEPQFLPDPH 341

Search completed: November 8, 2002, 19:29:22
Job time : 22.6993 secs

QY 133 YLLXIWP 140
Db 976 YLLSLWP 983

RESULT 13
US-08-424-788-6
; Sequence 6, Application US/08424788
; Patent No. 5716804
; GENERAL INFORMATION:
; APPLICANT: Moore, Kevin W.
; APPLICANT: Wei, Sherry
; APPLICANT: Ho, Alice Suk-Yue
; TITLE OF INVENTION: MAMMALIAN INTERLEUKIN-10 (IL-10)
; TITLE OF INVENTION: SUPER-ACTIVATING RECEPTORS; AND VARIANTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,788
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0501
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-424-788-6

Query Match 5.4%; Score 32; DB 1; Length 557;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
Db 307 LQTEPQFLLPDPH 320

RESULT 14
US-08-424-788-5
; Sequence 5, Application US/08424788
; Patent No. 5716804
; GENERAL INFORMATION:
; APPLICANT: Moore, Kevin W.
; APPLICANT: Wei, Sherry
; APPLICANT: Ho, Alice Suk-Yue
; TITLE OF INVENTION: MAMMALIAN INTERLEUKIN-10 (IL-10)
; TITLE OF INVENTION: SUPER-ACTIVATING RECEPTORS; AND VARIANTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto

STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,788
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 578 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-424-788-5

Query Match 5.4%; Score 32; DB 1; Length 578;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
Db 328 LQTEPQFLLPDPH 341

RESULT 15
US-08-110-683-2
; Sequence 2, Application US/08110683
; Patent No. 5789192
; GENERAL INFORMATION:
; APPLICANT: Moore, Kevin W.
; APPLICANT: Liu, Ying
; APPLICANT: Ho, Alice Suk-Yue
; APPLICANT: Hsu, Di-Hwei
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Tan, Jimmy C.
; APPLICANT: Chou, Chuan-Chu
; TITLE OF INVENTION: Mammalian Receptors for Interleukin-10
; TITLE OF INVENTION: (IL-10)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,683
; FILING DATE: 23-AUG-1993
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/011,066
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090

APPLICANT: Williams, Lewis T.
APPLICANT: Molz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,049A
FILING DATE: 29-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0637000US
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-609-049A-28

Query Match 5.6%; Score 33; DB 2; Length 1876;
Best Local Similarity 62.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 133 YLLXIWPX 140
||| :| :|
Db - 976 YLLSLWPP 983
RESULT 11
US-09-170-996-12
Sequence 12, Application US/09170996
Patent No. 6291220
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Molz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,996
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/609,049

FILING DATE: 29-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0637000US
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-12
Query Match 5.6%; Score 33; DB 4; Length 1876;
Best Local Similarity 62.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 133 YLLXIWPX 140
||| :| :|
Db 976 YLLSLWPP 983
RESULT 12
US-09-170-996-28
Sequence 28, Application US/09170996
Patent No. 6291220
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Molz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,996
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/609,049
FILING DATE: 29-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0637000US
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-28
Query Match 5.6%; Score 33; DB 4; Length 1876;
Best Local Similarity 62.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

US-09-036-987A-22

Query Match 5.8%; Score 34; DB 4; Length 278;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122

| | | | |

Db 159 LNTQQVAILPTSH 172

RESULT 4

US-09-370-700-22
Sequence 22, Application US/09370700
Patent No. 6274350

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J
APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22

LENGTH: 278

TYPE: PRT

ORGANISM: Saccharopolyspora spinosa

US-09-370-700-22

Query Match 5.8%; Score 34; DB 4; Length 278;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122

| | | | |

Db 159 LNTQQVAILPTSH 172

RESULT 5

US-09-071-035-76
Sequence 76, Application US/09071035
Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369p2
TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 76:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-76

Query Match 5.6%; Score 33; DB 4; Length 100;
Best Local Similarity 41.7%; Pred. No. 17;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 129 DXXXYLLXIWXP 140

| | | | |

Db 41 DTNRYLADVWVP 52

RESULT 6

US-09-071-035-74

Sequence 74, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369p2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 119 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-74

Query Match 5.6%; Score 33; DB 4; Length 119;
Best Local Similarity 41.7%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 129 DXXXYLLXIWXP 140

COUNTRY: USA
ZIP: 60603-2828
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette
COMPUTER: Compaq Prolinea 5100
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685.466C
FILING DATE: 24-JUL-96
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,381
FILING DATE: 24-JUL-95
ATTORNEY/AGENT INFORMATION:
NAME: Daniel J. Hulseberg
REGISTRATION NUMBER: 36,554
REFERENCE/DOCKET NUMBER: 96303755
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 701-8606
TELEFAX: (312) 701-7711
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Modified amino acid, formylmethionine
LOCATION: The first amino acid in the sequence,
LOCATION: described as methionine, is actually known as
LOCATION: formylmethionine ("fMet"). Polypeptide chains in
LOCATION: bacteria often start with formylmethionine.
PUBLICATION INFORMATION:
AUTHORS: L. de Boer, W. Harder, L. Dijkhuizen
TITLE: Phenylalanine and tyrosine metabolism in the
TITLE: facultative methyloleotroph No. 6015705ardia sp. 239
JOURNAL: Arch. Microbiol.
VOLUME: 149
PAGES: 459-465
DATE: 1988
PUBLICATION INFORMATION:
AUTHORS: J. Cairns, J. Overbaugh, S. Miller
TITLE:
JOURNAL: Nature
VOLUME: 335
PAGES: 142-145
DATE: 1988
PUBLICATION INFORMATION:
AUTHORS: G.J.W. Euverink, G.I. Hessels, J.W.
AUTHORS: Vrijbloed, J.R. Coggins, L. Dijkhuizen
TITLE: Purification and characterization of a dual
TITLE: function 3-hydroquinone dehydratase from Amycolatopsis
TITLE: methanolica
JOURNAL: J. Gen. Microbiology
VOLUME: 138
PAGES: 2449-2457
DATE: 1992
PUBLICATION INFORMATION:
AUTHORS: G.J.W. Euverink, D.J. Wolters, L.
AUTHORS: Dijkhuizen
TITLE: Prephenate dehydratase of the actinomycete
TITLE: Amycolatopsis methanolica: purification and
TITLE: characterization of the wild-type and deregulated mutant
TITLE: enzymes
JOURNAL: Biochem. J.
VOLUME: 308
PAGES: 313-320
DATE: 1995
PUBLICATION INFORMATION:
AUTHORS: P. Moretti, G. Hintermann, R. H tter
TITLE: Isolation and characterization of an
TITLE: extrachromosomal element from No. 6015705ardia mediterranei
JOURNAL: Plasmid

VOLUME: 14
PAGES: 126-133
DATE: 1985
PUBLICATION INFORMATION:
AUTHORS: J. Sambrook, E.F. Fritsch, T. Maniatis
TITLE: Molecular cloning: a laboratory manual
JOURNAL: Cold Spring Harbor Laboratory Press
DATE: 1989
PUBLICATION INFORMATION:
AUTHORS: J.W. Vrijbloed, J. Madon, L. Dijkhuizen
TITLE: A plasmid from the methyloleotrophic
TITLE: actinomycete Amycolatopsis methanolica capable of site-
TITLE: specific integration
JOURNAL: J. Bacteriol.
VOLUME: 176
PAGES: 7087-7090
DATE: 1994
US-08-685-466C-4
Query Match 5.8%; Score 34; DB 3; Length 192;
Best Local Similarity 24.1%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
QY 110 RTEHVXXLPDXHXXLXXXXXXLLXIW 138
DB 47 RFEHTVWPPVMAIVFNDTADHVLMAW 75
RESULT 3
US-09-036-987A-22
; Sequence 22, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

THIS PAGE BLANK (USPTO)

; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217

; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088722
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908

Query Match 5.3%; Score 31; DB 12; Length 1894;

Best Local Similarity 19.0%; Pred. No. 3.2e+02;

Matches 4; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 155 CXXXXXXXCCCCXXC 175

Db 557 CTTAACAAATAGCATACATC 577

Search completed: November 8, 2002, 19:29:44

Job time : 12.0654 secs

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 559 amino acids
TYPE: amino acid
STRADEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 496120
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-836-561-6

Query Match 5.3%: Score 31; DB 10; Length 559;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPD 120
||:| | | |
Db 59 LRVDHSVTLSD 70

RESULT 14

US-09-833-790-429
; Sequence 429, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833.790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-790-429

Query Match 5.3%: Score 31; DB 10; Length 629;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPD 120
||:| | | |
Db 395 LRSEESVSLPE 396

RESULT 15

US-10-052-586-97
; Sequence 97, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052.586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12406
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12406

Query Match 5.3%; Score 31; DB 10; Length 513;
Best Local Similarity 29.4%; Pred. No. 84;
Matches 5; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 104 KXXXXLRTHEHXXVXXLPD 120
| : : | | | |
DB 338 KTAQTIKSHHNVGGLPE 354

RESULT 11
US-09-815-242-10987
; Sequence 10987, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10987
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10987

Query Match 5.3%; Score 31; DB 10; Length 523;
Best Local Similarity 41.7%; Pred. No. 86;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 109 LRTEHXXVXXLPD 120
: : : | | | | |
DB 353 IKSHHNVGGLPD 364

RESULT 12
US-09-854-549-7
; Sequence 7, Application US/09854549
; Patent No. US20010034334A1
; GENERAL INFORMATION:
; APPLICANT: Ni, et al.
; TITLE OF INVENTION: Human Extracellular Matrix-1
; FILE REFERENCE: PF223C1
; CURRENT APPLICATION NUMBER: US/09/854,549
; CURRENT FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 09/007,105
; PRIOR FILING DATE: 1998-01-14
; PRIOR APPLICATION NUMBER: US 60/035,711
; PRIOR FILING DATE: 1997-01-16
; PRIOR APPLICATION NUMBER: US 60/050,113
; PRIOR FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-854-549-7

Query Match 5.3%; Score 31; DB 10; Length 559;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 109 LRTEHXXVXXLPD 120
| : : | | | | |
DB 59 LRVDHVSLSLHD 70

RESULT 13
US-09-836-561-6
; Sequence 6, Application US/09836561
; Patent No. US20020038006A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Corley, Neil C.
; Guegler, Karl J.
; TITLE OF INVENTION: HUMAN EXTRACELLULAR MATRIX PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/836,561
; FILING DATE: 16-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/212,168
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0333 US
; TELECOMMUNICATION INFORMATION:

; APPLICANT: Schaffner, Gotthold
 ; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
 ; FILE REFERENCE: 0652.1800001
 ; CURRENT APPLICATION NUMBER: US/09/970,711
 ; CURRENT FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/171.461
 ; PRIOR FILING DATE: 1999-01-12
 ; PRIOR APPLICATION NUMBER: PCT/EP97/01944
 ; PRIOR FILING DATE: 1997-04-18
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 13
 ; LENGTH: 206
 ; TYPE: PRT
 ; ORGANISM: CELO Virus
 ; FEATURE:
 ; OTHER INFORMATION: Position: 21134..21754 /gene: L3 /product: L3
 ; OTHER INFORMATION: protease
 US-09-970-711-13

Query Match 5.3%; Score 31; DB 10; Length 206;
 Best Local Similarity 42.9%; Pred. No. 33;
 ~Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
 || | : || |
 DB 192 LRRETALNALPENH 205

RESULT 8
 US-09-815-242-5490
 ; Sequence 5490, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5490
 ; LENGTH: 506
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5490

Query Match 5.3%; Score 31; DB 10; Length 506;
 Best Local Similarity 29.4%; Pred. No. 83;
 Matches 5; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 104 KXXXLRTEHXVXXLPD 120
 | : : | | |
 DB 331 KTAQTIKSHHNVGGLPE 347
 RESULT 9
 US-09-815-242-11555
 ; Sequence 11555, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11555
 ; LENGTH: 508
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori
 US-09-815-242-11555
 Query Match 5.3%; Score 31; DB 10; Length 508;
 Best Local Similarity 41.7%; Pred. No. 83;
 Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 109 LRTEHXVXXLPD 120
 : : | | |
 DB 338 IKTHHNVGGLPE 349
 RESULT 10
 US-09-815-242-12406
 ; Sequence 12406, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 42872
;; LENGTH: 30
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL031651.32
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
;; OTHER INFORMATION: EST_HUMAN HIT: BF529379.1, EVALUATE 3.00e-03
US-09-864-761-42872

Query Match 5.3%; Score 31; DB 10; Length 30;
Best Local Similarity 42.9%; Pred. No. 4.6;
Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
||:|:|:|:|
Db 17 LRPKHLAQPLPPAH 30

RESULT 6
US-09-864-761-34686
;; Sequence 34686, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecmica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 34686
;; LENGTH: 40
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP000459.2
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 73
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 28
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 83
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 19
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 59
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.3
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 24
;; OTHER INFORMATION: EXPRESSED IN HEILA, SIGNAL = 56
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 30
;; OTHER INFORMATION: EST_HUMAN HIT: BE300193.1, EVALUATE 4.00e-08
;; OTHER INFORMATION: SWISSPROT HIT: Q49412, EVALUATE 4.70e+00
US-09-864-761-34686

Query Match 5.3%; Score 31; DB 10; Length 40;
Best Local Similarity 35.7%; Pred. No. 6.2;
Matches 5; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
:|:|:|:|
Db 1 IHTNHFIILLKDNH 14

RESULT 7
US-09-970-711-13
;; Sequence 13, Application US/09970711
;; Patent No. US20020081279A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Adam
;; APPLICANT: Cotten, Matthew
;; APPLICANT: Chioocca, Susanna
;; APPLICANT: Kurzbauer, Robert

```
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 237
; LENGTH: 3503
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-237

Query Match          6.3%; Score 37; DB 9; Length 3503;
Best Local Similarity 33.3%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 120 DXHXXLXXDXHXXLLXIXWP 140
| | | | | | | | | | | | | | | |
DB 388 DGHFALTTRDNSIYLVHLP 408

RESULT 3
US-09-864-761-46893
; Sequence 46893, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
```

```
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46893
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL118499.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96
; OTHER INFORMATION: EST_HUMAN HIT: BE278123.1, EVALUE 1.00e-20
US-09-864-761-46893

Query Match          5.4%; Score 32; DB 10; Length 54;
Best Local Similarity 28.0%; Pred. No. 5.1;
Matches 7; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 113 HXVXXLPDXHXXLLXDXHXXLLXI 137
| | | | | | | | | | | | | | | |
DB 16 HVTATMPDLYHLLVELNAVQSLLGL 40

RESULT 4
US-09-946-143-2
; Sequence 2, Application US/09946143
; Patent No. US20020048795A1
; GENERAL INFORMATION:
; APPLICANT: FARWICK, Mike
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THECCSB GENE
; FILE REFERENCE: 032301 WD 212
; CURRENT APPLICATION NUMBER: US/09/946,143
; CURRENT FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-946-143-2

Query Match          5.4%; Score 32; DB 10; Length 337;
Best Local Similarity 62.5%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 133 YLLXIXWP 140
| | | | |
DB 210 YLLRWQP 217

RESULT 5
US-09-864-761-42872
; Sequence 42872, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:26:04 ; Search time 9.06536 Seconds
(without alignments)
464.428 Million cell updates/sec

Title: US-09-840-743-73
Perfect score: 590
Sequence: 1 FXXEXPXPXXXXXXX.....XXXXXXXXXXHXXSK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues
Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37	6.3	233	10 US-09-764-855-121	Sequence 121, App
2	37	6.3	3503	9 US-10-108-605-237	Sequence 237, App
3	32	5.4	54	10 US-09-864-761-46893	Sequence 46893, A
4	32	5.4	337	10 US-09-946-143-2	Sequence 2, Appli
5	31	5.3	30	10 US-09-864-761-42872	Sequence 42872, A
6	31	5.3	40	10 US-09-864-761-34686	Sequence 34686, A
7	31	5.3	206	10 US-09-970-711-13	Sequence 13, Appl
8	31	5.3	506	10 US-09-815-242-5490	Sequence 5490, Ap
9	31	5.3	508	10 US-09-815-242-11555	Sequence 11555, A
10	31	5.3	513	10 US-09-815-242-12406	Sequence 12406, A
11	31	5.3	523	10 US-09-815-242-10987	Sequence 10987, A
12	31	5.3	559	10 US-09-854-549-7	Sequence 7, Appli
13	31	5.3	559	10 US-09-836-561-6	Sequence 6, Appli
14	31	5.3	629	10 US-09-833-790-429	Sequence 429, App
15	31	5.3	1894	12 US-10-052-586-97	Sequence 97, Appl
16	30	5.1	141	10 US-09-799-777-16	Sequence 16, Appl
17	30	5.1	145	10 US-09-903-456-41	Sequence 41, Appl
18	30	5.1	148	10 US-09-903-456-51	Sequence 51, Appl
19	30	5.1	199	10 US-09-864-761-33578	Sequence 33578, A

20	30	5.1	199	10 US-09-864-761-33661	Sequence 33661, A
21	30	5.1	209	10 US-09-764-864-800	Sequence 800, App
22	30	5.1	210	10 US-09-811-284-229	Sequence 229, App
23	30	5.1	214	12 US-10-078-929-66	Sequence 66, Appl
24	30	5.1	234	10 US-09-912-020-350	Sequence 350, App
25	30	5.1	273	10 US-09-216-393-12	Sequence 12, Appl
26	30	5.1	283	10 US-09-903-456-85	Sequence 85, Appl
27	30	5.1	292	10 US-09-764-864-1261	Sequence 1261, Ap
28	30	5.1	292	10 US-09-903-456-65	Sequence 65, Appl
29	30	5.1	423	10 US-09-925-300-1233	Sequence 1233, Ap
30	30	5.1	502	10 US-09-323-988D-56	Sequence 56, Appl
31	30	5.1	1422	8 US-08-424-550B-81	Sequence 81, Appl
32	30	5.1	1497	10 US-09-060-854B-2	Sequence 2, Appli
33	30	5.1	4440	12 US-10-052-586-255	Sequence 525, App
34	30	5.1	4679	10 US-09-804-898-2	Sequence 2, Appli
35	29.5	5.0	272	10 US-09-925-118-2	Sequence 2, Appli
36	29.5	5.0	1386	10 US-09-866-582-38	Sequence 38, Appl
37	29	4.9	16	10 US-09-822-698A-51	Sequence 51, Appl
38	29	4.9	44	10 US-09-864-761-37419	Sequence 37419, A
39	29	4.9	60	9 US-10-046-961-17	Sequence 17, Appl
40	29	4.9	70	10 US-09-864-761-49072	Sequence 49072, A
41	29	4.9	80	10 US-09-894-018-97	Sequence 97, Appl
42	29	4.9	98	10 US-09-894-018-101	Sequence 101, App
43	29	4.9	106	10 US-09-894-018-95	Sequence 95, Appl
44	29	4.9	107	10 US-09-894-018-93	Sequence 93, Appl
45	29	4.9	107	10 US-09-894-018-103	Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-764-855-121
; Sequence 121, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 121
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (152)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (231)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-855-121

Query Match 6.3%; Score 37; DB 10; Length 233;
Best Local Similarity 46.2%; Pred. No. 1.8;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 110 RTEHVXXLPDXH 122
| | | : | | : |
Db 42 ROEHFIKGLPEYH 54

RESULT 2
US-10-108-605-237
; Sequence 237, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn

THIS PAGE BLANK (USPTO)

Best Local Similarity 26.7%; Pred. No. 26;
Matches 8; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

QY 111 TEHXVXXLPDXHXLXXDXXYYLXIIXWP 140
| : : | | | : | : |
Db 260 TQGMVNVNDSLVQLTPSDNPYYCLVLMVP 289

RESULT 15
F81423
L-lactate permease Cj0076c [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: F81423
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: F81423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-560 <PAR>
A:Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72562.1; PID:g696757
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: lctP; Cj0076c
C:Superfamily: L-lactate permease

Query Match 5.9%; Score 35; DB 2; Length 560;
Best Local Similarity 25.0%; Pred. No. 35;
Matches 8; Conservative 2; Mismatches 22; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXHXLXXDXXYYLXIIXWP 140
| : : | | | : | : |
Db 243 LSSNHLGAELPDIIISAVVSLAVTTVFLKEWKP 274

Search completed: November 8, 2002, 19:26:48
Job time : 24.085 secs

A:Residues: 1-1057 <STO>
A:Cross-references: GB:AE004723; GB:AE004091; NID:g9949067; PIDN:AAG06364.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: rne; PA2976
C:Superfamily: ribonuclease E

Query Match 6.1%; Score 36; DB 2; Length 1057;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
| : | : | | |
Db 461 LRTRARIFLPDDH 474

RESULT 12
A49335
1-phosphatidylinositol 4-kinase (EC 2.7.1.67) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein N0795; protein YNL267w
C:Species: Saccharomyces cerevisiae
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 29-Oct-1999
C:Accession: A49335; S44456; S60915; S63240; S65117; S39245
R:Flanagan, C.A.; Schnieders, E.A.; Emerick, A.W.; Kunisawa, R.; Admon, A.; Thorner, J.
Science 262, 1444-1448, 1993
A:Title: Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability
A:Reference number: A49335; MUID:94069320; PMID:8248783
A:Accession: A49335
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1066 <FA>
A:Cross-references: GB:L20220; NID:g432495; PIDN:AAA34873.1; PID:g432496
A:Accession: B49335
A:Molecule type: protein
A:Residues: 26-43;152-169;255-267;292-305;332-335;434-452;515-530;543-561;603-609;698-711
R:Garcia-Bustos, J.F.; Marini, F.; Stevenson, I.; Frei, C.; Hall, M.N.
EMBO J. 13, 2352-2361, 1994
A:Title: PIK1, an essential phosphatidylinositol 4-kinase associated with the yeast nucleus
A:Reference number: S44456; MUID:94252322; PMID:8194527
A:Accession: S44456
A:Molecule type: DNA
A:Residues: 1-1066 <GAR>
A:Cross-references: EMBL:X76058; NID:g433916; PIDN:CAA53658.1; PID:g433917
R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.H.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV from *Saccharomyces cerevisiae*
A:Reference number: S60909
A:Accession: S60915
A:Molecule type: DNA
A:Residues: 1-1066 <SEN>
A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63231.1; PID:g1045243
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63235
A:Accession: S63240
A:Molecule type: DNA
A:Residues: 1-1066 <SEM>
A:Cross-references: EMBL:Z71543; NID:g1302325; PIDN:CAA96174.1; PID:e239719; PID:g130232
A:Experimental source: strain S288C
R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.
Yeast 12, 505-514, 1996
A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from *Saccharomyces cerevisiae*
A:Reference number: S65111; MUID:96310631; PMID:8740425
A:Accession: S65117
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 721-960 <SEF>
A:Cross-references: EMBL:X92494
C:Genetics:
A:Gene: SGD; PIK1; PIK41; PIK120
A:Cross-references: SGD:S0005211; MIPS:YNL267w
A:Map position: 14L
C:Keywords: nucleus; phosphotransferase; transmembrane protein

F:156-172/Domain: transmembrane #status predicted <TMM>
Query Match 6.1%; Score 36; DB 2; Length 1066;
Best Local Similarity 35.7%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
| : | : | | |
Db 358 LNSDHTSSMPDLH 371

RESULT 13
D69786
glycoprotein endopeptidase homolog yd1c - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: D69786
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschl, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
isch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: D69786
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-229 <KUN>
A:Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12411.1; PID:g26329
A:Experimental source: strain 168
C:Genetics:
A:Gene: yd1c
C:Superfamily: hypothetical protein HI0388

Query Match 5.9%; Score 35; DB 2; Length 229;
Best Local Similarity 33.3%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXHXXDXD 129
| : | : | | | | |
Db 31 LRKNHVRAMPVHSLNDCD 51

RESULT 14
AB0013
probable exported protein YPO0099 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0013
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0013
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC88964.1; PID:gi5978208; GSPDB:GN00175
C:Genetics:
A:Gene: YPO0099

Query Match 5.9%; Score 35; DB 2; Length 411;

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C/Accession: B69305
C/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, G.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Pyrococcus furiosus strain ATCC 49239
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: B69305
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-239 <GLE>
A:Cross-references: GB:AE001074; GB:AE000782; NID:g2689397; PIDN:AAB90792.1; PID:g26525
C:Superfamily: conserved hypothetical protein H1191

Query Match 6.1%; Score 36; DB 2; Length 239;
Best Local Similarity 50.0%; Pred. No. 9.7;
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 129 DXXVYLKIXXP 140
| | | | |
DB 141 DTAVLANIWTTP 152

RESULT 10
AH0853
DNA mismatch repair protein [imported] - Salmonella enterica subsp. enterica serovar C/Species: Salmonella enterica subsp. enterica serovar typhi
A>Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C/Accession: AH0853
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, Th. T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Anatum
A:Reference number: AB0502; PMID:11677608
A:Accession: AH0853
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD06015.1; PID:g16503984; GSPDB:GN00176
C/Genetics:
A:Gene: SNY3033
C:Superfamily: DNA mismatch repair protein mutS

Query Match 6.1%; Score 36; DB 2; Length 855;
Best Local Similarity 35.0%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 110 RTEHXVXXLPDXHXLXXXD 129
| | | | |
DB 369 RMRAFOQLPELHAQLETVD 388

RESULT 11
H83273
ribonuclease E PA2976 [imported] - Pseudomonas aeruginosa (strain PAO1)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: H83273
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83273
A>Status: preliminary
A:Molecule type: DNA

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: D84781

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1207 <STO>

A:Cross-references: GB:AE002093; NID:94581149; PIDN:AAD24633.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g36490

A:Map position: 2

Query Match 15.4%; Score 91; DB 2; Length 1207;

Best Local Similarity 29.8%; Pred. No. 2.3e-11;

Matches 28; Conservative 0; Mismatches 56; Indels 10; Gaps 3;

QY 104 KXXXLRTXHXVXXLPDXHXXLX-----DXXXYLXIWXPPXXXXXXXXXXXXXX 157

DB 1094 KNISQLRTEHRYVLPDEHPLLAQLEKREPDPDCSYLLAIWTP-GETADSIOPSVSTC-- 1150

QY 158 XXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 191

DB 1151 -IFQANGMLCDEETCFSCNSIKETRQIVRGITL 1183

RESULT 3

T05430

hypothetical protein F28A23.180 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999

C:Accession: T05430

R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.; N

submitted to the Protein Sequence Database, October 1998

A:Reference number: Z15415

A:Accession: T05430

A:Molecule type: DNA

A:Residues: 1-917 <BEV>

A:Cross-references: EMBL:AL021961

A:Experimental source: cultivar Columbia; BAC clone F28A23

C:Genetics:

A:Map position: 4

A:Introns: 41/3; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/2; 624/3; 655/2; 667/3; 695

A:Note: F28A23.180

C:Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match

Best Local Similarity 13.1%; Score 77; DB 2; Length 917;

Matches 19; Conservative 0; Mismatches 16; Indels 6; Gaps 1;

QY 104 KXXXLRTXHXVXXLPDXHXXL-----XXDXXXYLLXIWXP 140

DB 836 KYNRLRTEHVVVLPDNDHLLHDFERRKLDPSYLLAIWQP 878

RESULT 4

S54723

UDP-glucose-glycoprotein glucosyltransferase (EC 2.4.1.-) precursor - fruit fly (Drosoph

C:Species: Drosophila melanogaster

C:Date: 23-Aug-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000

C:Accession: S54723; S54724

R:Parker, C.G.; Fessler, L.I.; Nelson, R.E.; Fessler, J.H.

EMBO J. 14, 1294-1303, 1995

A:Title: Drosophila UDP-glucose:glycoprotein glucosyltransferase: sequence and character

A:Reference number: S54723; MUID:95246722; PMID:7729408

A:Accession: S54723

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1548 <PAR>

A:Cross-references: GB:U20554; NID:g790584; PIDN:AAA85850.1; PID:g790585

A:Accession: S54724

A:Status: preliminary

A:Molecule type: protein

A:Residues: 23-37 <PA2>

C:Genetics:

A:Gene: FlyBase:Ugt

A:Cross-references: FlyBase:FBgn0014075

C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-1548/Product: UDP-glucose-glycoprotein glucosyltransferase #status experimental

F:181,266,864/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.9%; Score 41; DB 2; Length 1548;

Best Local Similarity 28.2%; Pred. No. 3.9;

Matches 11; Conservative 2; Mismatches 26; Indels 0; Gaps 0;

QY 96 PXXXXXXKXXLXTEHXXVXXLPDXHXXLXXXXXXVXL 134

DB 935 PQTKTRFKLPDLKTDHSHVVKLPKPKQLPHFDVAAML 973

RESULT 5

S73013

polyketide synthase pksc - Mycobacterium leprae

N:Alternate names: L518_F1_2 protein

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S73013

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid L518.

A:Reference number: S72591

A:Accession: S73013

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1446 <SMI>

A:Cross-references: EMBL:U00023; NID:9467194; PIDN:AAA17356.1; PID:9467199

C:Genetics:

A:Gene: pksc

C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-ox

homology; [acyl-carrier-protein] S-malonyltransferase homology

C:Keywords: carrier protein

F:54-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

F:563-840/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

F:1155-1334/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match

Best Local Similarity 6.4%; Score 38; DB 2; Length 1446;

Matches 8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 112 EHXVXXLPDXHXXLXXXDXXXY 133

DB 990 DHDVSTLHDLHNLGRGVNDVLY 1011

RESULT 6

H87203

polyketide synthase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 24-May-2001

C:Accession: H87203

R:Colle, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;

K.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: H87203

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1540 <STO>

A:Cross-references: GB:AL450380; NID:gl3093965; PIDN:CAC31872.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML2356

C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-ox

homology; [acyl-carrier-protein] S-malonyltransferase homology

C:Keywords: carrier protein

CiteSpace v. 5.8.R1
R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.: Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999

THIS PAGE BLANK (USPTO)

```

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH
CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA' CHAIN AND 1
CC OMEGA CHAIN (BY SIMILARITY).
CC -!- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
CC CORE RNAP. WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
CC TRANSCRIPTIONAL REGULATORS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003951; AAF83986.1; -
DR HSSP; P00574; ICGO.
DR InterPro; IPR001700; RNA_pol_A_bac.
DR Pfam; PF01000; RNA_pol_A_bac; 1.
DR Pfam; PF03118; RNA_pol_A_CTD; 1.
DR ProDom; PD001179; RNA_pol_A_bac; 1.
KW Transferase; Transcription; DNA-directed RNA polymerase;
KW -Complete proteome.
SQ SEQUENCE 332 AA; 36554 MW; 98BDE535F753A577 CRC64;

Query Match 5.6%; Score 33; DB 1; Length 332;
Best Local Similarity 35.7%; Pred. No. 17;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 109 LRTEHXXVXXLPDXH 122
Db 115 IKTDHNVETINDGH 128
:::| | : | |

```

Search completed: November 8, 2002, 19:27:16
Job time : 15.4967 secs


```

DR TIGRFAMS; TIGR01080; rplX_A_E; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein.
SQ SEQUENCE 142 AA; 16075 MW; B496362F0BFEC354 CRC64;

Query Match
Best Local Similarity 5.6%; Score 33; DB 1; Length 142;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 109 LRTEHVVXXLPD 119
      ||||| : :|
DB 38 LRTKHGIRAIP 48

RESULT 13
TDXH_PVRAB STANDARD; PRT; 216 AA.
AC Q9U2V4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable peroxiredoxin.
GN PAB1673.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN# [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. TDXH SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AP000005; BAA30317.1; -.
DR HSSP; P30041; 1PRX.
DR InterPro; IPR000866; AhpC-TSA.
DR Pfam; PF00578; AhpC-TSA; 1.
KW Antioxidant; Complete proteome.
FT ACT_SITE 46 46 BY SIMILARITY.
SQ SEQUENCE 216 AA; 24757 MW; A95DF14BF3874CA9 CRC64;

Query Match
Best Local Similarity 5.6%; Score 33; DB 1; Length 216;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 109 LRTEHVVXXLPD 120
      :||| | |||
DB 13 VKTHTGVKLPD 24

RESULT 15
RPOA_XYLFA STANDARD; PRT; 332 AA.
AC Q9PE52;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase
DE alpha chain) (RNA polymerase alpha subunit).
DN RPOA OR XF1176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

```

```
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 314 334 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 440 460 POTENTIAL.
FT TRANSMEM 538 558 POTENTIAL.
SQ SEQUENCE 560 AA; 58920 MW; 8ECB9F626FFC40A CRC64;

Query Match 5.8%; Score 34; DB 1; Length 560;
Best Local Similarity 30.4%; Pred. No. 17; Mismatches 15; Indels 0; Gaps 0;
Matches 7; Conservative 1;

Oy 118 LPDXHXXLXXXDXLXIXWP 140
||| | | | |
Db 249 LPDITSALVSIVSLALFLKVRP 271

RESULT 11
CYAA_NEUCR
ID CYAA_NEUCR STANDARD; PRT; 2300 AA.
AC Q01631;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl
cyclase).
DE CR-1 OR NAC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP MEDLINE=92000795; PubMed=1680356;
RX Kore-Eda S., Murayama T., Uno I.;
RT "Isolation and characterization of the adenylate cyclase structural
gene of Neurospora crassa.";
RL Jpn. J. Genet. 66:317-334(1991).
CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR
METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,
CAMP.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-3 FAMILY.
CC -1- SIMILARITY: CONTAINS 23 LEUCINE-RICH REPEATS (LRR).
CC -1- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00909; BAA00755.1;
DR InterPro; IPR001054; G.cyclase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR InterPro; IPR001932; PP2C-like.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00211; quanylate_cyc; 1.
DR Pfam; PF00481; PP2C; 1.
DR Pfam; PF00560; LRR; 14.
DR Pfam; PF00788; RA; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00044; CYC; 1.
DR SMART; SM00370; LRR; 4.
DR SMART; SM00369; LRR_TYP; 2.
DR SMART; SM00332; PP2Cc; 1.
```

```
DR SMART; SM00314; RA; 1.
DR PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.
KW Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.
LRR 1.
FT REPEAT 867 890 LRR 1.
FT REPEAT 892 914 LRR 2.
FT REPEAT 915 938 LRR 3.
FT REPEAT 939 961 LRR 4.
FT REPEAT 962 986 LRR 5.
FT REPEAT 988 1008 LRR 6.
FT REPEAT 1009 1031 LRR 7.
FT REPEAT 1033 1055 LRR 8.
FT REPEAT 1056 1079 LRR 9.
FT REPEAT 1081 1097 LRR 10.
FT REPEAT 1098 1120 LRR 11.
FT REPEAT 1122 1142 LRR 12.
FT REPEAT 1143 1165 LRR 13.
FT REPEAT 1166 1188 LRR 14.
FT REPEAT 1189 1211 LRR 15.
FT REPEAT 1213 1234 LRR 16.
FT REPEAT 1234 1369 LRR 17.
FT REPEAT 1373 1396 LRR 18.
FT REPEAT 1398 1420 LRR 19.
FT REPEAT 1422 1445 LRR 20.
FT REPEAT 1447 1469 LRR 21.
FT REPEAT 1474 1497 LRR 22.
FT REPEAT 1513 1538 LRR 23.
FT DOMAIN 1564 1829 PP2C-LIKE.
FT DOMAIN 1830 2300 CATALYTIC.
FT DOMAIN 49 52 POLY-SER.
FT DOMAIN 204 221 POLY-GLN.
SQ SEQUENCE 2300 AA; 254752 MW; 52E79B90B6B17A7B CRC64;

Query Match 5.8%; Score 34; DB 1; Length 2300;
Best Local Similarity 31.6%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Oy 104 KXXXXLRTXHXVXXLPDXH 122
| | | | |
Db 462 KTSSSARSGHSHVLPQGH 480

RESULT 12
RL26_CAEEL
ID RL26_CAEEL STANDARD; PRT; 142 AA.
AC Q19869;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L26.
GN RPL-26 OR F28C6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Burton J.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z68315; CAA92674.1;
DR WormPep; F28C6.7; CE03278.
DR InterPro; IPR000302; KOW_motif.
DR Pfam; PF00467; KOW; 1.
```



```
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; AAC59088.1; -.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Zinc-finger.
FT ZN_FING 8 66 RING-TYPE.
SQ SEQUENCE 249 AA; 28295 MW; 03B17D50BA9EE4B CRC64;

Query Match 5.8%; Score 34; DB 1; Length 249;
Best Local Similarity 50.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTHEVXXLPDXH 122
DB 193 LRAERAVKALADAH 206

RESULT 7
TDXH_AERPE
ID TDXH_AERPE STANDARD; PRT; 250 AA.
AC QYV9L0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable peroxiredoxin.
GN APE2278.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. TDXH SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000064; BAA81290.1; -.
DR HSSP; P30041; IPRX.
DR InterPro; IPR000866; AhpC-TSA.
DR Pfam; PF00578; AhpC-TSA; 1.
KW Antioxidant; Complete proteome.
FT ACT_SITE 50 50 BY SIMILARITY.
SQ SEQUENCE 250 AA; 28703 MW; 0457F285D051E7A CRC64;
```

```
Query Match 5.8%; Score 34; DB 1; Length 250;
Best Local Similarity 50.0%; Pred. No. 7.5;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 111 TEHXVXXLPDXH 122
DB 19 TDHGVIKLPDHY 30

RESULT 8
MURF_TREPA
ID MURF_TREPA STANDARD; PRT; 459 AA.
AC O83401;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAC-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
GN MURF OR TP0386.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete";
RL Science 281:375-388(1998).
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine - ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001217; AAC65370.1; -.
DR HSSP; P11880; 1GG4.
DR TIGR; TP0386;
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR Pfam; PF01225; Mur_ligase; 1.
DR Pfam; PF02875; Mur_ligase_C; 1.
DR TIGRFAMS; TIGR01143; murF; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW ATP-binding; Complete proteome.
FT NP_BIND 121 127 ATP (POTENTIAL).
SQ SEQUENCE 459 AA; 50234 MW; 23EA5571ED95F332 CRC64;

Query Match 5.8%; Score 34; DB 1; Length 459;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
CC      step. This protein has a weak ATPase activity.
CC      -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC      -!- CAUTION: Ref.1 sequence differs from that shown due to
CC      frameshifts in positions 353, 379, 394, 422 and 425.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U16303; AAA08578.1; ALT_FRAME.
DR      EMBL; M18965; AAA27167.1; ALT_FRAME.
DR      EMBL; AE008832; AAL21789.1; -.
DR      EMBL; AL627276; CAD06015.1; -.
DR      PIR; A28668; A28668.
DR      StyGene; SG10237; mutS.
DR      InterPro; IPR000432; Muts_C.
DR      InterPro; IPR002863; Muts_N.
DR      Pfam; PF00488; Muts_C; 1.
DR      Pfam; PF01624; Muts_N; 1.
DR      SMART; SM00534; MutsSac; 1.
DR      TIGRFAMS; TIGR01070; mutS1; 1.
DR      PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
DR      DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT      NP_BIND 616 623      ATP (POTENTIAL).
FT      MUTAGEN 622 622      K->A: DEFECTIVE IN MISMATCH REPAIR;
FT      ATPASE ACTIVITY REDUCED 6 FOLD.
FT      TA -> S (IN REF. 1).
DR      CONFLICT 361 362
DR      CONFLICT 855 AA; 95406 MW; 804BEB13BCD2F6F CRC64;
SQ      SEQUENCE 855 AA; 95406 MW; 804BEB13BCD2F6F CRC64;

Query Match      6.1%; Score 36; DB 1; Length 855;
Best Local Similarity 35.0%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY      110 RTEHXVXXLPDXHXXLXXXD 129
      | | | | | | | |
DB      369 RMRHAFQQLPELHAQLETVD 388

RESULT 5
PIKL_YEAST
ID      PIKL_YEAST      STANDARD;      PRT; 1066 AA.
AC      P39104;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Phosphatidylinositol 4-kinase PIK1 (EC 2.7.1.67) (PI4-kinase)
DE      (PtdIns-4-kinase).
GN      PIK1 OR YNL267W OR N0795.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JK9-3D;
RX      MEDLINE=94252322; PubMed=8194527;
RA      Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N.;
RT      "PIK1, an essential phosphatidylinositol 4-kinase associated with the
RT      yeast nucleus.";
RL      EMBO J. 13:2352-2361(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94069320; PubMed=8248783;
RA      Flanagan C.A., Schmieders E.A., Emerick A.W., Kunisawa R., Admon A.,
RT      Thorne J.;
RT      "Phosphatidylinositol 4-kinase: gene structure and requirement for
RT      yeast cell viability.";
RL      Science 262:1444-1448(1993).
RN      [3]
```

```
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / FY1679;
RX      MEDLINE=96310631; PubMed=8740425;
RA      Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
RT      "The sequence of a 24,152 bp segment from the left arm of chromosome
RT      XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT      genes.";
RL      Yeast 12:505-514(1996).
CC      -!- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST
CC      COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER
CC      INOSITOL-1,4,5-TRISPHOSPHATE. PIK1 IS PART OF A NUCLEAR
CC      PHOSPHOINOSITIDE CYCLE AND COULD CONTROL CYTOKINESIS THROUGH
CC      THE ACTIN CYTOSKELETON.
CC      -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-ID-myo-inositol = ADP +
CC      1-phosphatidyl-ID-myo-inositol 4-phosphate.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X76058; CAA53658.1; -.
DR      EMBL; L20220; AAA34873.1; -.
DR      EMBL; X92494; CAA63231.1; -.
DR      EMBL; Z71543; CAA96174.1; -.
DR      PIR; A49335; A49335.
DR      PIR; S44456; S44456.
DR      SGD; S0005211; PIK1.
DR      InterPro; IPR000403; PI3_PI4_Kinase.
DR      Pfam; PF00454; PI3_PI4_kinase; 1.
DR      SMART; SM00146; PI3Kc; 1.
DR      PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR      PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR      PROSITE; PS02090; PI3_4_KINASE_3; 1.
DR      TRANSFERASE; Kinase; Nuclear protein.
FT      DOMAIN 793 1041      PI3K/PI4K.
SQ      SEQUENCE 1066 AA; 119922 MW; 7666979CA14B1CB5 CRC64;

Query Match      6.1%; Score 36; DB 1; Length 1066;
Best Local Similarity 35.7%; Pred. No. 12;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      109 LRTEHXVXXLPDXH 122
      | : | | | |
DB      358 LNSDHTSSMPDLH 371

RESULT 6
VCG3_NPVOP
ID      VCG3_NPVOP      STANDARD;      PRT; 249 AA.
AC      O10339;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Zinc-finger protein CG30.
GN      CG30.
OS      Orgyia pseudotsugata multicausid polyhedrosis virus (OpMPNV).
OC      Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC      Nucleopolyhedrovirus.
OX      NCBI_TaxID=164623;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97271300; PubMed=9126251;
RA      Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA      Rohrmann G.F.;
RT      "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT      polyhedrosis virus genome.";
RL      Virology 229:381-399(1997).
```



```
KW Glycoprotein. 1 22
FT SIGNAL 23 1548
FT CHAIN
FT
FT CARBOHYD 181 181.
FT CARBOHYD 266 266.
FT CARBOHYD 864 864.
FT SITE 1545 1548
SQ SEQUENCE 1548 AA; 174465 MW; 95D6849961622DB6 CRC64;

Query Match 6.9%; Score 41; DB 1; Length 1548;
Best Local Similarity 28.2%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 11; Conservative 2; Mismatches 26;

QY 96 PXXXXXXXLTHXVXLPDXHXXLXXXXXXXL 134
DB 935 PQTQTRFKLPDLDKTHSVVVKLPKQKLPHEFDVAVL 973

RESULT 2
ID DYHC_SCHPO STANDARD; PRT; 4196 AA.
AC O13290; Q9P6L0; Q9UTP8;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Dynein heavy chain, cytosolic (DYHC).
GN DHC1 OR SPAC30C2.01C OR SPAC1093.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CRL152;
RA Yamamoto A., West R.R., McIntosh J.R., Hiraoka Y.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouris J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Bozzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Goux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RT Nature 415:871-880(2002).
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES. REQUIRED FOR NUCLEAR MOVEMENT
CC DURING MEIOTIC PROPHASE.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:14 ; Search time 10.4967 Seconds
(without alignments)
1153.796 Million cell updates/sec

Title: US-09-840-743-73

Perfect score: 590

Sequence: 1 PXEXPPXPXXXXXXX.....XXXXXXXXXXLHXXSK 292

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	41	6.9	1548	1	UGGG_DROME	Q09332 drosophila
2	38	6.4	1496	1	DYHC_SCHPO	Q13290 schizosacch
3	37	6.3	1593	1	ATL2_HUMAN	P58397 homo sapien
4	36	6.1	855	1	MOTS_SALTY	P10339 salmonella
5	36	6.1	1066	1	PIK1_YEAST	P39104 saccharomyc
6	34	5.8	249	1	VCG3_NPVOP	O10339 orgyia pseu
7	34	5.8	250	1	TDXH_AERPE	Q9y910 aeropyrum p
8	34	5.8	459	1	MURE_TREPA	O83401 treponema p
9	34	5.8	510	1	GUAA_AQUAE	O66601 aquifex aeo
10	34	5.8	560	1	GLCA_ECOLI	Q46839 escherichia
11	34	5.8	2300	1	CYAA_NEUCR	Q01631 neurospora
12	33	5.6	142	1	RL26_CAEEL	Q19869 caenorhabdi
13	33	5.6	216	1	TDXH_PYRAB	Q9uzv4 pyrococcus
14	33	5.6	216	1	TDXH_PYRHO	O58966 pyrococcus
15	33	5.6	332	1	RPOA_XYLFA	Q3pe22 xyliella fas
16	33	5.6	355	1	DVRI_BRARE	P35621 brachydanio
17	33	5.6	380	1	T701_FREDI	Q00461 fremyella d
18	33	5.6	521	1	DRTS_TRYCR	Q27793 trypanosoma
19	33	5.6	1023	1	TSCC_PSEAM	P55019 pseudopleur
20	33	5.6	1098	1	RP16_DROME	Q39qno drosophila
21	33	5.6	1275	1	RFBC_MYXXA	Q50864 myxococcus
22	33	5.6	3924	1	ANK2_HUMAN	Q01484 homo sapien
23	32.5	5.5	293	1	TYRC_ZYMMO	Q04983 zymomonas m
24	32.5	5.5	517	1	SEST_CAEEL	Q9n4d6 caenorhabdi
25	32	5.4	190	1	RHOL_DROME	Q24192 drosophila
26	32	5.4	232	1	RFAY_SALTY	P26472 salmonella
27	32	5.4	430	1	MESJ_HABIN	P44689 haemophilus
28	32	5.4	490	1	YE0J_SCHPO	Q10990 schizosacch
29	32	5.4	492	1	WZXC_ECOLI	P77377 escherichia
30	32	5.4	537	1	CNEI_HUMAN	Q99829 homo sapien
31	32	5.4	555	1	HEXA_CAEEL	Q22492 caenorhabdi
32	32	5.4	578	1	IL0R_HUMAN	Q13651 homo sapien
33	32	5.4	775	1	PFL_STRMU	Q59934 streptococc

ALIGNMENTS

RESULT 1

UGGG_DROME	5-4	787	1	PFL_LACLA	032797 lactococcus
ID	UGGG_DROME	STANDARD;	PRT;	1548 AA.	
AC	Q09332;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	UDP-glucose:glycoprotein glucosyltransferase precursor (EC 2.4.1.-)				
DE	(UDP-Glc:glycoprotein glucosyltransferase) (dUGT).				
GN	UGT OR UGGG.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;				
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-37.				
RC	TISSUE=Embryo;				
RX	MEDLINE=95246722; PubMed=7729408;				
RA	Parker C.G., Fessler L.I., Nelson R.E., Fessler J.H.;				
RT	"Drosophila UDP-glucose:glycoprotein glucosyltransferase: sequence				
RT	and characterization of an enzyme that distinguishes between				
RT	denatured and native proteins."				
RL	EMBO J. 14:1294-1303(1995).				
CC	-!- FUNCTION: UNFOLDED, DENATURED GLYCOPROTEINS ARE SUBSTANTIALLY				
CC	BETTER SUBSTRATES FOR GLYCOSYLATION BY THIS ENZYME THAN ARE THE				
CC	CORRESPONDING NATIVE PROTEINS. THIS PROTEIN AND TRANSIENT				
CC	GLYCOSYLATION MAY BE INVOLVED IN MONITORING AND/OR ASSISTING THE				
CC	FOLDING AND ASSEMBLY OF NEWLY MADE GLYCOPROTEINS, IN ORDER TO				
CC	IDENTIFY GLYCOPROTEINS THAT NEED ASSISTANCE IN FOLDING FROM				
CC	CHAPERONES.				
CC	-!- COFACTOR: REQUIRES CALCIUM AND MANGANESE IONS FOR ACTIVITY.				
CC	-!- PATHWAY: Glycosylation.				
CC	-!- SUBUNIT: MONOMER.				
CC	-!- SUBCELLULAR LOCATION: Endoplasmic reticulum.				
CC	-!- DEVELOPMENTAL STAGE: IS PRESENT AT LOW BUT DETECTABLE LEVELS IN				
CC	THE EARLIEST EMBRYOS, INCREASING AT 6-8 HRS WITH A MAXIMUM AT 10-				
CC	12 HRS. LEVELS DECREASE THEREAFTER AND ARE NOT DETECTED IN 18-20				
CC	HRS EMBRYOS AND FIRST INSTAR LARVAE BUT IS DETECTED AGAIN AT				
CC	SECOND INSTAR TO PUPATION.				
CC	-!- SIMILARITY: SOME, TO YEAST KRE5, S.TYPHIMURIUM RFAJ AND E.COLI				
CC	REAI PROTEINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed, usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U20554; AAA85850.1; -				
DR	FlyBase; FBgn0014075; Ugt.				
DR	InterPro; IPR002495; GT 8.				
DR	Pfam; PF01501; Glyco_transf_8; 1.				
KW	Signal; Transferase; Glycosyltransferase; Endoplasmic reticulum;				

THIS PAGE BLANK (USPTO)

```

|||||
Db 301 TTGAACAGATAGTACAACTACTGGACATGAATCCAGAGCGAGATCTGACAAAAGTA 360
QY 1480 tgcagagcattatgactcgtctgtctgttaatgcagcgaagtaactatgcagcaaaatgatg 1539
|||||
Db 361 TCGAGAGCAATTATGGACTCGTCTGCTGTTAAATGCGACGAAGCTACTGAACAAAATGATG 420
QY 1540 gcagcagacaagtgtctctggagttcgaccttaacaaaactcctcagcagaacccctcca 1599
|||||
Db 421 GCAGCAGACAAAGTGTCTGGAGTTCGACCTTAACAAAACCTCTCTGCGAAGCCCTCCA 480
QY 1600 aaagga 1605
|||||
Db 481 NAAGGA 486

RESULT 7
AV545791/c
LOCUS AV545791 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION AV545791 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
CDNA clone RZL01d12F 3', mRNA sequence.
ACCESSION AV545791
VERSION AV545791.1 GI:8717205
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 575)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..575
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12F"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 167 a 116 c 119 g 173 t
ORIGIN

Query Match 6.9%; Score 477; DB 9; Length 575;
Best Local Similarity 99.8%; Pred. No. 4.7e-68;
Matches 488; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6365 gagaggaagtttccgtcaacgggacatatctccaagtaacagagattattgcagacca 6424
Db 575 GAGAGGAAGTTTCCGTCACAGGGACATATTCCAAGTCAACGAGTATTTCGACACCA 516
QY 6425 cgagtcaggtctcaaacccatcgatgtctctcctagattggatctggatctcccaagaag 6484
Db 515 CGAGTCAGTCTCAACCCATCGATGTTCTAGAGATTGGATATGGATCTCCCAAGAG 456
QY 6485 gactgtttacttcggaacatcagtaacataatctcagaggtctttcaacggagcagat 6544
Db 455 GACTGTTTACTTCGGAACATCAGTAACATCAATATTTCAGAGGCTTTTCAACGGAGCAGAT 396
QY 6545 acagttctgcttttgaaagattcgtatgttcctggttcgattcgaacagacaagc 6604
|||||

```

```

Db 395 ACAGTCTCTCTTTTGGAAAGGATTCTGTATGTCTCGTGGATTGGAACAGAAAGACAGAGC 336
QY 6605 accggtccatttaattgcaaggttgctatttctcctgcgagcaaatggaagcaacaaac 6664
|||||
Db 335 ACCGGTCCATTAAATGGCAAGCTTGCAATTTCTCGAGCAAAATGGAACACAAAC 276
QY 6665 ctaagatgactcgggaagaaagcaaacgcattctctctctctctctctctctctctctct 6724
Db 275 CTAAGATGACTGGGAAGAAAGCAAAACGCAATTCCTCTCTCTCTCTCTCTCTCTCTCT 216
QY 6725 ggaagagtcctatttagacataataacaggaatcaaataggctatttctctctctctct 6784
Db 215 GGAAGAGTCCCATTTAGACATAATAACAGGAATCAATAGGCTATTTTCTCTCTCTCT 156
QY 6785 ttatttcattcatagcagagcgacac-aaaaagtttttgggttatttatttcttc 6843
|||||
Db 155 TTATTTTCATCATAGCAGAGCGACACAAAAAAGTTTTCGGTTATTATTATTTCTC 96
QY 6844 tctacaaa 6852
|||||
Db 95 TCTAACAA 87

RESULT 8
BG123849
LOCUS EST469495 tomato shoot/meristem Lycopersicon esculentum cDNA clone
DEFINITION cTOF3G24 5' sequence, mRNA sequence.
ACCESSION BG123849
VERSION BG123849.1 GI:12624037
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 759)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Ronning,C. and Tanksley,S.
Generation of ESTs from tomato shoot/meristem tissue
Unpublished (2001)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES
source
1..759
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOF3G24"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 225 a 158 c 191 g 185 t
ORIGIN

Query Match 6.0%; Score 410; DB 10; Length 759;
Best Local Similarity 73.4%; Pred. No. 3.5e-57;
Matches 524; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 5062 tcagaaaagatgtggagggaatgaaggagagagcaagcaacaaacaaatattgatt 5121
Db 4 TGAGAAAGGAAGTCCCAATCAAGAGTGGGAAAAAGAAAGCAAGGATGCAATGAGT 63

```



```

Source 1..623
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="R2L05c12P"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 177 a 127 c 135 g 184 t
ORIGIN

Query Match 7.7%; Score 528.4; DB 9; Length 623;
Best Local Similarity 99.6%; Pred. No. 1.9e-76;
Matches 540; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6312 gcaaacacacagagcttcaggaactctctgataccttgggactgcatgagagga 6371
Db 623 GCAAACTCAGAGACAGTTCGAGAACTCTCTGATACCTTGTGGACTGCCATGAGAGGA 564
QY 6372 agtttcgcgtcaacggagacatattccaaagtcaacagagtatttgagaccagagtc 6431
Db 563 AGTTTCCGCTCAACGGGACATATTCCAAAGTCAACGAGTTATTGACAGACCAGAGTCC 504
QY 6432 agtctcaaacacacagcttcctagagattggatggatctcccaagaagagactgtt 6491
Db 503 AGTCTCAAAACCCATCGATGTTCTAGAGATTGGATATGGGATCTCCCAAGAGGACTGT 444
QY 6492 tacttcggaacatcagtaacataattcagaggtcttccacggagcagatcacagttc 6551
Db 443 TACTTCGGAACATCATAGTAACATCAATATTTCAGAGGTCTTTCAACGGAGCAGATACAGTTC 384
QY 6552 tgctttggaaaggattcgatgtccgtgattcgaaagacagaaagacacgcgt 6611
Db 383 TGCTTTGAAAGGATTCGATGTGTCGTGATTTCGAAACAGAAAGACGACCGCGT 324
QY 6612 ccattaatggcaaggtgctatttccgtcgagcaaatgaaagacacacacacacacac 6671
Db 323 CCATTAAATGGCAAGTTCGATTTCCTCGGACCAATTTGAAGAACACAAACCTTAAAGA 264
QY 6672 tgactggaagaaagcaaacgactgtctctctctctctctctctctctctctctct 6731
Db 263 TGACTGGAAGAAAGCAACGCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 204
QY 6732 tccatttagacataataacaggaatccaaatagagctatttctctctctctctctct 6791
Db 203 TCCATTATAGACATATAACAGGAATCCAAATAGGCTATTCTCTCTCTCTCTCTCTCT 144
QY 6792 attcatagacagagagacac-aaaaagtttttgggtatttatttctctctctct 6850
Db 143 ATTCATAGACAGAGCGACACAAAAAAGTTTTTGGGTATTATTATCTCTCTCTAACA 84
QY 6851 aa 6852
Db 83 AA 82

RESULT 4
AV556668
LOCUS
DEFINITION
AV556668 Arabidopsis thaliana green silique Columbia Arabidopsis
ACCESSION
AV556668
VERSION
AV556668.1 GI:8728083
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 527)
REFERENCE
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
AUTHORS

A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..527
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0049d09f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 160 a 113 c 112 g 142 t
ORIGIN

Query Match 7.6%; Score 525.4; DB 9; Length 527;
Best Local Similarity 99.8%; Pred. No. 6.2e-76;
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6276 gacgagactgttctgagtgtaacagctgagggagcaaacctcacagacaggttcgagga 6335
Db 1 GACGAGACTGTCTTGAGTGTAACTGCTGAGGAGCAAACTCACAGACAGTTCGAGGA 60
QY 6336 actcttcgataccttctggactgccatgagaggaagtttccgctcaacggacatat 6395
Db 61 ACTCTTCTGATACCTTGTGGACTGCCATGAGAGGAAGTTTTCGCTCAACGGGACATAT 120
QY 6396 ttccaagtcaacgagttatttgcagaccacagagtcacagtcctcaaacccatgatgtct 6455
Db 121 TTCCAAGTCAACGAGTTATTTCGACAGCAGCAGTCCAGTCTCAAAACCCATCGATCTCT 180
QY 6456 agagattggatggatctcccaagaagagagctgtttacttcgaaacatcaatcaatca 6515
Db 181 AGAGATTGGATGGATGCTCCCAAGAGAGGACTGTTTACTTCGGAACATCAATCAATCA 240
QY 6516 atattcagaggttcttcaacggagcagatcacagttctgtcttttgaaaggattcgatgt 6575
Db 241 ATATTCAGAGGTCTTTCAACGGAGCAGATACAGTTCTGCTTTTGGAAAGGATTGCTATGT 300
QY 6576 gtccgtggattcgaacagagacagagcaccgctccatttaattggaaggttgcatttt 6635
Db 301 GTCCGTGGATTCCGAACAGAGACAAAGACCCGCTCATTAATGGAAGGTTGCAATTTT 360
QY 6636 cctgcgagcaaatgaagaacacaaacccctaaagatgactggaagaagcaaacgactt 6695
Db 361 CCTGCGAGCAAAATTGAAGACACAAACCTTAAGATGACTGGAAAGAAACCAACGCATT 420
QY 6696 gcttctgtctctctctctatttaagccaggaagttccatttagacataataacagga 6755
Db 421 GCTTCTGTGCTCTCTCTATTATAAGCCAGGAAAGTTCCATTATAGACATATAACAGGA 480
QY 6756 atccaaataggtctatttctctctctctctctctctctctctctctctctctctct 6802
Db 481 ATCCAAATAGGCTATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 527

RESULT 5
B60854/c
LOCUS
DEFINITION
B60854 T1907TF TAMU Arabidopsis thaliana genomic clone T1907, DNA
sequence.
ACCESSION
B60854
VERSION
B60854.1 GI:2627785
KEYWORDS
GSS.

```

The 11st of best scores is:

130 140 150
PLPXXLTQLHLEXPXXXXXQKXLMPR
|||:|||||;
PLPESTQLHLEL
550 X

The 11st of best scores is:

Result NO.	Query			DB	ID	Description
	Score	Match	Length			
1	228	32.9	555	2	T48453	hypothetical prote
2	201.5	29.1	917	2	T05430	hypothetical prote
3	139	20.1	1207	2	D84781	hypothetical prote
4	112.5	16.2	259	2	D25735	endonuclease III -
5	89.5	12.9	236	2	S75373	probable DNA-(apur
6	89.5	12.9	265	2	D84198	endonuclease III f
7	86.5	12.5	241	2	G87530	endonuclease III f
8	86	12.4	222	2	A75109	endonuclease III (
9	83.5	12.0	233	2	B69202	endonuclease III -
10	82	11.8	222	2	F71025	probable, endonucle
11	78.5	11.3	338	2	F75459	probable endonucle
12	76	11.0	232	2	F72383	hypothetical prote
13	74.5	10.8	572	2	T52520	hypothetical prote
14	74	10.7	218	2	G86758	DNA-(apurinic or a
15	72	10.4	219	2	AF1311	probable endonucle
16	72	10.4	219	2	AF1683	probable endonucle
17	70	10.1	214	2	B71919	endonuclease III -
18	70	10.1	218	2	A64593	endonuclease III -
19	69.5	10.0	213	2	H70325	endonuclease III -
20	69	10.0	218	2	B83862	endonuclease III (
21	68.5	9.9	264	2	C72770	probable DNA-(apur
22	68	9.8	219	2	D89923	endonuclease-like
23	67	9.7	209	2	C69461	endonuclease III (
24	64	9.2	219	2	T40525	endonuclease III (
25	63.5	9.0	212	2	F71634	endonuclease III (
26	62.5	9.0	225	2	B75537	endonuclease III -
27	62	8.9	297	2	C87149	probable DNA glyco
28	61.5	8.9	292	2	B97233	8-oxoguanine-DNA-g
29	61	8.8	209	2	F95148	endonuclease III -

F28A23.180 - Arabidopsis thaliana
 Species: Arabidopsis thaliana (mouse-ear cress)
 Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_c
 Accession: T05430
 Submitted to the Protein Sequence Database, October 1998
 Reference number: Z15415
 Accession: T05430
 Molecule type: DNA
 Residues: 1-917 <BEV>
 Cross-references: EMBL:AL021961
 Experimental source: cultivar Columbia; PAC clone F28A23.180

A; Note: F28A23.180

> 0 <
01 10 Intelligenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file seq72-q941x6.res made by bobryen on Tue 17 Dec 102 12:58:40 PST.

Query sequence being compared: SEQ72-Q941X6 (1-132)
Number of sequences searched: 5
Number of scores above cutoff: 5

Results of the initial comparison of SEQ72-Q941X6 (1-132) with:
File: seq72compares.pep

```

100-
N -
U -
M -
B -
E -
R -
O -
F -
S -
E -
Q -
U -
E -
C -
N -
E -
S -
SCORE 0 11 22 33 44 55 67 78 89 100
STDDEV -1 0 0 0 0 0 0 0 0 0

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 132
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 69 Median 30 Standard Deviation 36.96
Times: CPU 00:00:00.90 Total Elapsed 00:00:01.00

Number of residues: 3817
Number of sequences searched: 5
Number of scores above cutoff: 5

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Opt. Score	Sig. Frame
1. q9sr66	T22K18.18 protein.	1309	100	121	0.84 0
2. t48453	TOIG of: t48453 check: 2431	555	96	119	0.73 0
3. seq72-q941x6	Putative FPP synthase (Fragment)	119	93	115	0.65 0

1. SEQ72-Q941X6 (1-132)
q9sr66 T22K18.18 protein.

Initial Score = 100 Optimized Score = 121 Significance = 0.84
Residue Identity = 56% Matches = 75 Mismatches = 57
Gaps = 1 Conservative Substitutions = 0

```

SKESAKSTQKKSVDMSYRKKEAESGGRKRETERPTDTPDMLRCTDVAKIANLITIRGNMNLAEIKAF
780 790 800 810 820 830 840
LXXXVXXHGXIDLEMLRXXPPDAXKXLLSTXGLKRSVECVRLIHXKXEPVDTN-GRIVRLGVPLOP
60 70 80 90 100 110 120
LNRLVKKHSGSDLEMLRDPDPKAKETLSINGLKSVECVRLLSHIAFPVDTN-GRIVRLGVPLOP
850 860 870 880 890 900 910 920
130 X
LPXXXQXHX
LPDELQMHLLERLELHYHMITFGKRVFCTK
930 940

```

2. SEQ72-Q941X6 (1-132)
t48453 TOIG of: t48453 check: 2431 from: 1 to: 555
Initial Score = 96 Optimized Score = 119 Significance = 0.73
Residue Identity = 54% Matches = 73 Mismatches = 58
Gaps = 1 Conservative Substitutions = 1

```

MKGTLDGKPTSDMSLRKDVEGNEGROENFTNMDSIDYEALRRASLSISAIERGNMNLAEIKAF
400 410 420 430 440 450 460 470
LXXXVXXHGXIDLEMLRXXPPDAXKXLLSTXGLKRSVECVRLIHXKXEPVDTN-GRIVRLGVPLOP
60 70 80 90 100 110 120
LNRLVKKHSGSDLEMLRDPDPKAKETLSINGLKSVECVRLLSHIAFPVDTN-GRIVRLGVPLOP
480 490 500 510 520 530 540 550
LERIVKHGGIDLEMLRSPDPKAKDVLISIRGLKRSVECVRLIHXKXEPVDTN-GRIVRLGVPLOP

```

130 X
LPXXXQXHX
LPESLQHLLEL
550 X

3. SEQ72-Q941X6 (1-132)
seq72-q941x6 Putative FPP synthase (Fragment).

Initial Score = 93 Optimized Score = 115 Significance = 0.65
Residue Identity = 60% Matches = 72 Mismatches = 46
Gaps = 1 Conservative Substitutions = 0

```

WXXXRXVXXHGXIDLEMLRXXPPDAXKXLLSTXGLKRSVECVRLIHXKXEPVDTN-GRIVRLGVPLOP
X 10 20 30 40 50 60 70
LGMCIEMOVQDDYLDMEAVRCADVDIELANTIKERGNMNLAEIKAF
X 10 20 30 40 50 60

```



DB 882 LGLKSVCEVRLLSHQIAFPVDTNNGRIAVRLGWPLQPLPDLQHLL----- 930

QY 157 XLWPRCKLXQXTLVELHXXITFGKXXFCTXXXPNCNACPMXXECXXXSA 209

DB 931 -----ELVELHYHMITFGK-VFCTKVKNCNACPM-KAECRHYSSA 969

RESULT 2

Q9LZ68

ID Q9LZ68 PRELIMINARY; PRT; 555 AA.

AC Q9LZ68

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE Hypothetical 62.1 kDa protein.

GN T32M21-170.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,

RA De Clerck R., De Keyser A., Nuyt P., Rouze P., Van Den Daele H.,

RA Villaroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,

RA Rud S., Lemcke K., Mayer K.F.X.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL162875; CAB85563.1;

DR InterPro; IPR003265; Endo_3c.

DR Pfam; PF00730; Hhh-GPD; 1.

DR SMART; SM00478; ENDO3G; 1.

DR Hypothetical protein.

QY SEQUENCE 555 AA; 62088 MW; 14B44BDEE17FDC1E CRC64;

Query Match 32.9%; Score 228; DB 10; Length 555;

Best Local Similarity 44.5%; Pred. No. 1.6e-29;

Matches 57; Conservative 0; Mismatches 61; Indels 10; Gaps 1;

QY 18 DXXXXXXXXXXXXXXIXRGXXXXXXLRXXXXXXXFLXXXXXXGIDLE 77

DB 436 DSIDYEAIRRASTISEISEAIEKRGMMMLAVRI-----KDFLERIVKHGIDLE 485

QY 78 WLRXXXXXXXKXLLXXXXGKLSXCEVRLRLXXXXXAFVDTNNGRIAVRXGXVPLXPLP 137

DB 486 WUESPPDAKDYLLSIRGLKLSXCEVRLRLHNLAFVDTNNGRIAVRMGWVPLQPLP 545

QY 138 XXXQXHL 145

DB 546 ESLQLHL 553

RESULT 3

Q94LX6

ID Q94LX6 PRELIMINARY; PRT; 119 AA.

AC Q94LX6

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DE Putative FPPsynthase (Fragment).

GN ETPFPS1.

OS Euphorbia tirucalli.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids I; Malpighiales; Euphorbiaceae; Euphorbia.

NCBI_TaxID=142860;

RN [1]

RP SEQUENCE FROM N.A.

RA Asada T., Fukusaki E., Kobayashi A., Sando T.,

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB051105; BAB47587.1;

DR InterPro; IPR003265; Endo_3c.

DR Pfam; PF00730; Hhh-GPD; 1.

DR NON_TER 1

FT NON_TER 119

SQ SEQUENCE 119 AA; 13771 MW; FDBCE4CD0C5B81 CRC64;

Query Match 29.1%; Score 201.5; DB 10; Length 917;

Best Local Similarity 50.5%; Pred. No. 8.1e-26;

Matches 50; Conservative 0; Mismatches 39; Indels 10; Gaps 1;

QY 37 IXRGXXXXXXLRXXXXXXXFLXXXXXXGIDLEWLRXXXXXXXKXLLXXG 96

DB 31 IKRGNNVLAKEI-----KDFNLRLVREHGNIDLEWLRDVPFDRKAYLLSIRG 80

QY 97 XGLKXCEVRLRLXXXXXAFVDTNNGRIAVRXGXVPLXPLP 135

DB 81 LGLKSVCEVRLRLHNLAFVDTNNGRIAVRLGWVPLQPLP 119

RESULT 4

Q49498

ID Q49498 PRELIMINARY; PRT; 917 AA.

AC Q49498

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DE Hypothetical 106.3 kDa protein.

GN F28A23.180 OR AT4G34060.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,

RA Herzl A., Neumann S., Hoheisel J., Jesse T., Heijnen L., Vos P.,

RA Mewes H.W., Mayer K., Schueller C.,

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,

RA Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL021961; CAA17566.1;

DR EMBL; AL161584; CAB80123.1;

DR InterPro; IPR003265; Endo_3c.

DR InterPro; IPR003651; PEs_Bind.

DR Pfam; PF00730; Hhh-GPD; 1.

DR SMART; SM00478; ENDO3G; 1.

DR SMART; SM00525; FES; 1.

DR Hypothetical protein.

QY SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match 29.1%; Score 201.5; DB 10; Length 917;

Best Local Similarity 49.5%; Pred. No. 7.8e-25;

Matches 46; Conservative 0; Mismatches 22; Indels 25; Gaps 3;

QY 117 VDTNNGRIAVRXGXVPLXPLPXXXXXHXHXXLPXXXXXXKXLLXXQXTLVELHYX 176

DB 625 VDTNNGRIAVRLGWVPLPPLPNGVQMHQLE-----YELHYQ 661

QY 177 XITFGKXXFCTKXXPCNACPMXXECXXXSA 209

DB 662 MITFGK-VFCTKTIPNCNACPM-KSECKYFASA 692

RA Mewes H.W., Mayer K., Schueller C.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBDJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.
 DR EMBL; AL021961; CAAL7566.1; -
 DR EMBL; AL161584; CAB80123.1; -
 DR InterPro; IPR003265; Endo_3c.
 DR InterPro; IPR003651; FeS_Bind.
 DR Pfam; PF00730; RHH-GPD; 1.
 DR SMART; SM00478; ENDO3c; 1.
 DR SMART; SM00525; FES; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACEIA454 CRC64;
 Query Match 58.9%; Score 162; DB 10; Length 917;
 Best Local Similarity 43.9%; Pred. No. 1.3e-19;
 Matches 43; Conservative 0; Mismatches 47; Indels 8; Gaps 2;
 QY 1 KVVXDXXTXXXWXL-----XXDXDXXXXXXXERXXXFXRXXXGXGXXX 53
 DB 418 KVNLDPEIKEDVLMVNDSPSRSDYDKTEAKWKEREIFQTRIDLFNRHLOGNRK 477
 QY 54 FXXKSGVVDVXGVFLTON-XDXSSXAXMXXAXFP 90
 DB 478 FKQKSGVVDVXGVFLTONTTDYLSSNAFMSVAAKFP 515
 RESULT 4
 ID Q9LZ69 PRELIMINARY; PRT; 1017 AA.
 AC Q9LZ69;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE Hypothetical 112.1 kDa protein.
 GN T3M21.160.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
 RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
 RA Villarroel R., Gellen J., Van Montagu M., Bancroft I., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.
 DR EMBL; AL162875; CAB85562.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1017 AA; 112139 MW; 4F7C313A891EC83F CRC64;
 Query Match 18.0%; Score 49.5; DB 10; Length 1017;
 Best Local Similarity 28.8%; Pred. No. 9.4;
 Matches 17; Conservative 0; Mismatches 33; Indels 9; Gaps 1;
 QY 1 KVVXDXXTXXXWXL-----XXDXDXXXXXXXERXXXFXRXXXGXGXXX 50
 DB 931 KVDIDETTRWNLMLCKGDEKDKKKEKWEERVFGRGRADSFARMHLVQG 989

DB 557 TFSWKGVSVDVXGVFLTONVADHSSSAYWLAEEFP 595
 RESULT 2
 ID Q9SJQ6 PRELIMINARY; PRT; 1207 AA.
 AC Q9SJQ6;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE At2g36490 protein.
 GN AT2G36490.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Cleary T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.
 DR EMBL; AC006919; AAD24633.1; -
 DR InterPro; IPR003651; FeS_Bind.
 DR SMART; SM00525; FES; 1.
 SQ SEQUENCE 1207 AA; 135916 MW; AADB6C9EE94DDF67 CRC64;
 Query Match 59.5%; Score 163.5; DB 10; Length 1207;
 Best Local Similarity 43.4%; Pred. No. 9.7e-20;
 Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;
 QY 1 KVVXDXXTXXXWXL-----XXDXDXXXXXXXERXXXFXRXXXGXGXXX 52
 DB 500 KVDIDETTRWNLMLCKGDEKAKWEEERVFGRGRADSFARMHLVQGR 559
 QY 53 FXXKSGVVDVXGVFLTON-XDXSSXAXMXXAXFP 90
 DB 560 RFTWKGVSVDVXGVFLTONVSHLSSAFMSLASQFP 598
 RESULT 3
 ID O49498 PRELIMINARY; PRT; 917 AA.
 AC O49498;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Hypothetical 106.3 kDa protein.
 GN F28A23.180 OR AT4G34060.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
 RA Herzl A., Neumann S., Hoheisel J., Jesse T., Heijnen L., Vos P.,